INTERNATIONAL ATOMIC ENERGY AGENCY
UNITED NATIONS EDUCATIONAL, SCIENTIFIC AND CULTURAL ORGANIZATION
INTERNATIONAL CENTRE FOR THEORETICAL PHYSICS
I.C.T.P., P.O. BOX 586, 34100 TRIESTE, ITALY, CABLE CENTRATOM TRIESTE

SMR.572/2

SCHOOL ON DYNAMICAL SYSTEMS

(9 - 27 September 1991)

Homoclinic Bifurcations and Sensitive-Chaotic Dynamics

Jacob Pallis
Instituto de Matematica Pura e Aplicada (IMPA)
Edificio Lelio Gama
Estrada Dona Castorina 110
Jardim Botanico C.P. 34021
Rio de Janeiro 22460
Brazil

These are preliminary lecture notes, intended only for distribution to participants

ICTP

Summer School on Dynamical Systems
September, 1991

HOMOCLINIC BIFURCATIONS AND SENSITIVE-CHAOTIC DYNAMICS

Jacob Palis

Text: Jacob Palis and Floris Takens

CONTENTS

PREFACE

Homoclinic bifurcations, which form the main topic of this monograph, belong to the area of dynamical systems, the theory which describes mathematical models of time evolution, like differential equations and maps. Homoclinic evolutions, or orbits, are evolutions for which the state has the same limit both in the "infinite future" and in the "infinite past".

Such homoclinic evolutions, and the associated complexity, were discovered by Poincaré and described in his famous essay on the stability of the solar system around 1890. This associated complexity was intimately related with the break-down of power series methods, which came to many, and in particular to Poincaré, as a surprise.

The investigations were continued by Birkhoff who showed in '35 that in general there is near a homoclinic orbit an extremely intricate complex of periodic solutions, mostly with a very high period.

The theory up to this point was quite abstract: though the inspiration came from celestial mechanics, it was not proved that in the solar system homoclinic orbits actually can occur. Another development took place which was much more directed to the investigation of specific equations: in order to model vacuum tube radio receivers Van der Pol introduced in '20 a class of equations, now named after him, describing nonlinear oscillators, with or without forcing. His interest was mainly in the periodic solutions and their dependence on the forcing. In later investigations of this same type of equations, around '50, Cartwright, Littlewood and Levinson discovered solutions which were much more complicated than any solution of a differential equation known up to that time.

Now we can easily interpret this as complexity caused by (the suspension of) a horseshoe, which in its turn is a consequence of the existence of one (transverse) homoclinic orbit, but that is inverting the history...

In fact, Smale, who originally had focussed his efforts on gradient and gradient-like dynamical systems, realized, when confronted with these complexities, that he should extend the scope of his investigations. Seventy years after Poincaré, Smale was again shocked by the complexity of homoclinic behaviour! By the mid 60's he had a very simple geometric example (i.e. no formula's but just a picture and a geometric description), which could be completely analysed and which showed all the complexity found before: the horseshoe. This new prototype dynamical model, the horseshoe, together with the investigations in the behaviour of geodesic flows of manifolds with negative curvature (Hadamard, Anosov). grew, due to the efforts of a number of mathematicians, to an extension of the gradient-like theory which we now know as hyperbolic dynamics, and which in particular provides models for very complex (chaotic) dynamic behaviour.

Around '80 this hyperbolic theory was used by Levi to reanalyse the qualitative behaviour of the solutions of Van der Pol's equation, largely extending the earlier results. He proved that, besides all the complexity we know from the hyperbolic theory, even the new and extreme complexity associated with homoclinic bifurcations, which we shall consider below, actually exists in the solutions of this equation.

Homoclinic bifurcations, or non-transverse homoclinic orbits, become important when going beyond the hyperbolic theory. In the late 60's, Newhouse combined homoclinic bifurcations with the complexity already available in the hyperbolic theory to obtain dynamical systems far more complicated than the hyperbolic ones. Ultimately this led to his famous result on the coexistence of infinitely many periodic attractors and was also influential to our own work on hyperbolicity or lack of it near homoclinic bifurcations. These developments form the main topic of the present monograph, of which we shall now outline the content.

We start with a chapter presenting general background information about the hyperbolic theory and its relation to (structural) stability of systems, and discuss as well some initial aspects of chaotic dynamics - many results on stable manifolds and foliations are stated, and their proofs sketched, in Appendix I. The later chapters, except the last one, do not depend on the results described in Chapter 0 and are basically self-contained.

In Chapter I, we give a number of simple examples of homoclinic orbits and bifurcations. Chapter II discusses the horseshoe example and shows how it is related to homoclinic orbits. Then, in Chapter III, we consider some elementary consequences of the occurence

of a homoclinic bifurcation, especially in terms of cascades of bifurcations which have to accompany them.

In the Chapters IV, V, and VI, we come to our main topic: the investigation of situations where there is an interplay between homoclinic bifurcations and non-trivial basic sets, the sets being the building blocks of hyperbolic systems with complex behaviour. Since such basic sets have often a fractal structure, we start in Chapter IV with a discussion of Cantor sets and fractal dimensions like the Hausdorff dimension. In Chapter V the emphasis is on hyperbolicity near a homoclinic bifurcation associated with a basic set of small Hausdorff dimension. Then, in Chapter VI, we discuss types of homoclinic bifurcations which lead, in a persistent way, to complexity beyond hyperbolicity. In this chapter we provide a new, and more geometric, proof of Newhouse's result on the coexistence of infinitely many periodic attractors. Finally, in Chapter VII we present an overview of recent results, including specially Hénon-like strange attractors. We also pose new conjectures and problems which lead to a more complete picture of the place of homoclinic bifurcations in dynamics.

In our presentation we mainly restrict ourselves to diffeomorphisms in dimension two (which is the proper context to investigate e.g. the forced Van der Pol equation), although some extensions to higher dimensions are mentioned; also we concentrate mainly on the general theory as opposed to the analysis of specific equations. Consequently a number of topics like Silnikov's bifurcations and the Melnikov method are not discussed.

We hope that by puting this material together, rearranging it to some extent and pointing to recent and possible future directions, these results and their proofs will become more accessible, and will find their central place in dynamics which we think they merit.

We wish to thank a number of colleagues and Ph. D. students from the Instituto de Matematica Pura e Aplicada (IMPA) who much helped us in writing this book. Among them we mention M. Carvalho, L. Diaz, P. Duarte, R. Mañé, L. Mora, S. Newhouse, M. J. Pacífico, J. Rocha, D. Ruelle, R. Ures, J. C. Yoccoz and most specially M. Viana.

CHAPTER 0

HYPERBOLICITY, STABILITY AND SENSITIVE-CHAOTIC DYNAMICAL SYSTEMS

In this chapter we give background information and references to the literature concerning basic notions in dynamical systems that play an important role in our study of homoclinic bifurcations. Essentially, the chapter consists of a summary of the hyperbolic theory of dynamical systems and comments on sensitive chaotic dynamics. This is intended both as an introduction to the following chapters and to provide a more global context for the results to be discussed later and in much more detail than the ones presented in this 0-chapter.

In the first section we concentrate on hyperbolicity and emphasize its intimate relation with various forms of (structural) stability. In the second section we discuss several aspects of sensitivity ("chaos") and indicate how it occurs in hyperbolic systems.

§1. Hyperbolicity and stability

These two concepts, hyperbolicity and (structural) stability, have played an important role in the development of the theory of dynamical systems in the last decades: the hyperbolic theory was mostly developed in the sixties, having as a main initial motivation the construction of structurally stable systems; in its turn, the notion of structural stability had been introduced much earlier by Andronov and Pontryagin [**AP**, 1937]. As conjectured in the late sixties and only recently proved, it turns out that the two notions are *essentially equivalent* to each other, at least for C^1 diffeomorphisms of a compact manifold. *Of course, for stability one also has to impose the transversality of all stable and unstable manifolds or, for limit set-stability, the no-cycle condition; see concepts below.* It is, however, the hyperbolicity of the limit set which is the main ingredient in this comparison. The solution of this well known conjecture and other related results that we state here go beyond what is needed to understand the next chapters of this book. It is, however, enlightning, in the study of bifurcations (meaning loss of stability), to be acquainted with the fact that the notions of stability and hyperbolicity are that much interconnected.

The concept of (structural) stability deals with the topological persistence of the orbit structure of a dynamical system (endomorphism, diffeomorphism or flow) under small perturbations. (Notice the difference with the concept of Liapunov stability which concerns attracting sets of a given system). The persistence of the orbit structure is expressed in terms of a ***homeomorphism*** of the ambient manifold sending orbits of the initial system onto orbits of the perturbed one. If this can be done for any C^k small ($k \geq 1$) perturbation, then we call the system C^k ***(structurally) stable*** or ***globally stable***. Here we are mostly concerned with diffeomorphisms, in which case we require this orbit preserving homeomorphism to be a ***conjugacy***. That is, if φ is the initial map, $\tilde{\varphi}$ a C^k small perturbation of it, we then require the existence of a homeomorphism h such that $h\varphi(x) = \tilde{\varphi}h(x)$ for all x in the ambient manifold. We do not require the conjugacy to be differentiable, for otherwise we would impose invariance of eigenvalues of the linear part of the diffeomorphism at fixed (periodic) points. The same concept applies to endomorphisms, i.e. maps from the ambient manifold to itself.

We will be treating here the case of n-dimensional diffeomorphisms; for diffeomorphisms of the circle (and flows on surfaces) there are early important works of Pliss [**P**,1960], Arnold [**A**,1961] and specially Peixoto [**P**,1962]. Often we are concerned with stability restricted to the main part of the orbit structure, the limit set or nonwandering set. Let us recall these concepts.

Let $\varphi: M \to M$ be a C^k ($k \geq 1$) diffeomorphism of a compact, boundaryless, smooth manifold of arbitrary dimension. For $x \in M$, we define the α and ω limit sets as

$$\alpha(x) = \{y \in M; \exists n_i \to -\infty \text{ such that } \varphi^{n_i}(x) \to y\}$$
$$\omega(x) = \{y \in M; \exists n_i \to +\infty \text{ such that } \varphi^{n_i}(x) \to y\}.$$

The ***positive*** and ***negative limit sets*** are then defined as $L^+(\varphi) = \overline{\bigcup_{x\in M} \omega(x)}$ and $L^-(\varphi) =$

$\overline{\bigcup_{x\in M} \alpha(x)}$; the *limit set* $L(\varphi)$ is the union of $L^+(\varphi)$ and $L^-(\varphi)$. From the definitions, it is clear that $L^+(\varphi)$ and $L^-(\varphi)$ are φ-invariant, i.e. $\varphi(L^+(\varphi)) = L^+(\varphi)$ and $\varphi(L^-(\varphi)) = L^-(\varphi)$. Moreover, for each $x \in M$, $\varphi^n(x)$ approaches $L^+(\varphi)$ or $L^-(\varphi)$ as $n \to \infty$ or $n \to -\infty$. So $L^+(\varphi)$ and $\varphi|L^+(\varphi)$, $L^-(\varphi)$ and $\varphi|L^-(\varphi)$, describe the asymptotic behaviour of orbits, i.e. sequences $\{\varphi^n(x)\}$, in M for $n \to \infty$ or $n \to -\infty$. Another relevant concept is that of ***nonwandering*** point: x is nonwandering if for any neighbourhood U of it, there is an integer n such that $\varphi^n(U) \cap U \neq \phi$. Again, the union of the nonwandering points, which is denoted by $\Omega(\varphi)$, is a φ-invariant compact set. Clearly, all α or ω-limit points as well as homoclinic points (see Chapter I) are nonwandering. In Section 4, Chapter V, we provide an example of a homoclinic tangency which is in $L^+(\varphi)$ but not in $L^-(\varphi)$, or vice-versa; this example also shows that, in general, the nonwandering and limit sets are different. However, as we shall see in Chapter II, any transversal homoclinic point is accumulated by periodic orbits and so it is in $L^+(\varphi)$, in $L^-(\varphi)$, and in $\Omega(\varphi)$. Finally, we define another useful concept: the chain recurrent set, which is the union of the chain recurrent points. A point p is chain recurrent if for each $\epsilon > 0$ there are points $x_0 = p, x_1, x_2, \cdots, x_k = p$ such that $d(f(x_{i-1}), x_i) < \epsilon$ for $1 \leq i \leq k$, d being a distance function.

If $L^+(\varphi)$ (or $L^-(\varphi)$) is hyperbolic (see Chapter II and Appendix I), then one can show that $\overline{\text{Per}(\varphi)} = L^+(\varphi)$ (or $L^-(\varphi)$), where $\text{Per}(\varphi)$ indicates the set of periodic points of φ; one can then write as in **[N,1972]**:

$$L^+(\varphi) = \Lambda_1 \cup \cdots \cup \Lambda_k$$

where each Λ_i is ***invariant***, ***compact***, ***transitive*** (it has a dense orbit) and has a ***dense subset of periodic orbits***. This is called the spectral decomposition of $L^+(\varphi)$. Moreover, by **[HPPS, 1970]** (see also **[N,1980]**, **[B,1977]** for a different and relevant proof using the idea of "shadowing" of orbits), each Λ_i is the ***maximal invariant set*** in a neighbourhood of it. This last fact is actually equivalent to what we call ***local product structure*** in Λ_i: there exist $\epsilon > 0$ and $\delta > 0$ such that if the distance between $x, y \in \Lambda_i$ is smaller than δ then their local stable and unstable manifolds of size ϵ (see Appendix I) intersect each other in a unique point and this point is in Λ_i. Also, one can prove that if $\omega(x) \subset \Lambda_i$ then $x \in W^s(z)$ for some $z \in \Lambda_i$. In general, a set with the properties above is called a ***basic set*** for the diffeomorphism.

If we assume that the nonwandering set $\Omega(\varphi)$ is hyperbolic and $\overline{\text{Per}(\varphi)} = \Omega(\varphi)$, then we say that φ satisfies ***Axiom A***. In this case we have $\Omega(\varphi) = L^+(\varphi)$ and so we can write the nonwandering set as a finite union of basic sets. This is the content of Smale's spectral decomposition theorem **[S,1970]**; the corresponding version for the limit set as presented above appeared later in **[N,1972]**. Notice that if $\Lambda_1 \cup \cdots \cup \Lambda_k$ is the spectral decomposition of $L^+(\varphi)$ (or $\Omega(\varphi)$) then $M = \bigcup_i W^s(\Lambda_i)$, where $W^s(\Lambda_i) = \{y; \omega(y) \subset \Lambda_i\}$ is called the stable set of Λ_i; as discussed above $W^s(\Lambda_i) = \bigcup_{x\in\Lambda_i} W^s(x)$. Similar statements are valid for the unstable sets of the Λ_i's, corresponding to a spectral decomposition of $L^-(\varphi)$ or $\Omega(\varphi)$. Some $W^s(\Lambda_i)$ must be open; in this case Λ_i is called an ***attractor***. (A more general definition of attractor is in the next section). Dually if $W^u(\Lambda_i)$ is open, then we say that Λ_i is a ***repeller***. Finally, Λ_i is of ***saddle-type*** if it is neither an attractor nor a repeller. Another property of Axiom A diffeomorphisms: the stable sets of attractors cover an open and dense subset of M and the same is true for unstable sets of repellers. It is an interesting fact that if φ is C^2, then the union of the stable sets of attractors has total Lebesgue measure **[R,1976;BR,1975]**. There are, however, examples of C^1 saddle-type basic sets with stable sets of positive Lebesgue measure **[B,1975]**, which are detailed in Chapter IV. Another interesting fact about basic sets is that they are ***expansive***: for each basic set Λ there is a constant $\alpha > 0$ such that for each pair of different points in Λ, their (full) orbits get apart by at least α. ***From this it follows that hyperbolic attractors which are not just fixed or periodic sinks have sensitive dependence on initial conditions***: for most pairs of different points in the stable set of such an attractor Λ, the positive orbits get apart by at least a constant (which depends on the attractor). ***Most*** here means probability one in $W^s(\Lambda) \times W^s(\Lambda)$.

The following relevant result concerning basic sets states that they are persistent under C^k small perturbations (see Appendix I); in particular, hyperbolic attractors are persistent.

Theorem. *If Λ is a basic set for a C^k diffeomorphism $\varphi: M \to M$, then for any $\tilde{\varphi}$ close to φ its maximal invariant set in some neighbourhood of Λ is a basic set $\tilde{\Lambda}$ and $\varphi|\Lambda$ is conjugate to $\tilde{\varphi}|\tilde{\Lambda}$. Moreover, if we require the conjugacy to be C^0 close to the inclusion map of Λ into M, then it is unique and it is in fact Hölder continuous.*

Usually we call this set $\tilde{\Lambda}$ *the ("smooth") continuation of* Λ for a given perturbation of φ. The result can be applied when all of M is hyperbolic for φ; in this case we say that φ is Anosov. As a corollary, we have:

Theorem [A,1967]. *Anosov diffeomorphisms are globally C^k stable.*

Moser's elegant proof of this last result [M,1969] actually suggested the original proof of the first of the two previous theorems, but of course Anosov's theorem was proved before. Nowadays, there is a simple way of showing the existence of the conjugacy; one uses again the idea of shadowing of orbits mentioned above.

Another relevant class of systems in our context is that of Morse-Smale diffeomorphisms. We call φ Morse-Smale if

i) $\Omega(\varphi)$ consists of a finite number of periodic orbits, all of them hyperbolic;

ii) φ satisfies the transversality condition: the stable and unstable manifolds of the elements in $\Omega(\varphi)$ are all in general position.

It turns out that in (i) above one can write $L^+(\varphi)$ or $L^-(\varphi)$ instead of $\Omega(\varphi)$. *Morse-Smale diffeomorphisms are abundant in the sense that they contain the time-one maps of an open and dense subset of gradient vector fields on every manifold.*

Theorem [PS,1970]. *Morse-Smale diffeomorphisms are C^k stable. (In particular, there are stable diffeomorphisms on every manifold.) Conversely, among diffeomorphisms whose nonwandering sets consist of finitely many periodic orbits, hyperbolicity of these orbits and transversality of their stable and unstable manifolds are necessary for C^k stability.*

In view of these results, it seemed that hyperbolicity of the nonwandering set (or limit set) and transversality of stable and unstable manifolds were the precise conditions that should grant C^k stability of the diffeomorphism. So, let us say that an Axiom A

diffeomorphism f satisfies the transversality condition if the stable and unstable manifolds of any two points in $\Omega(f)$ are in general position.

Stability Conjecture [PS,1970]: A C^k (or $C^s, s \geq k$) diffeomorphisms is C^k stable if and only if it satisfies Axiom A and the transversality condition.

Remark: We can phrase the stability conjecture in terms of the limit set (or just the positive or negative limit set): a diffeomorphism is C^k stable if and only if its limit set is hyperbolic and the transversality condition holds. This equivalent, and perhaps more elegant, statement is further commented below.

Much in parallel, let us start discussing stability restricted to the nonwandering set or to the limit set. We say that a diffeomorphism φ is C^k Ω-stable if there exists a C^k neighbourhood of it such that if $\tilde{\varphi}$ belongs to this neighbourhood, then $\varphi|\Omega(\varphi)$ is conjugate to $\tilde{\varphi}|\Omega(\tilde{\varphi})$. Similarly for the limit set. Let now φ be an Axiom A diffeomorphism and let $\Omega(\varphi) = \Lambda_1 \cup \cdots \cup \Lambda_k$ be its spectral decomposition. A j-cycle on Ω is a string of j pairs of points $x_1, y_1 \in \Lambda_{i_1}, \cdots, x_j, y_j \in \Lambda_{i_j}$, with not all $i_1, \cdots, i_j$ equal, such that $W^u(y_1) \cap W^s(x_2) \neq \phi, \cdots, W^u(y_j) \cap W^s(x_1) \neq \phi$. If the limit set, or just the positive or negative limit set, is hyperbolic then we also have a spectral decomposition for it and the notion of cycles can be applied. In either case, when there are *no cycles* we can construct a *filtration* ([S,1967],[C,1978]): a sequence $M_0 = \phi$, $M_1 \subset M_2 \subset \cdots \subset M_k = M$ of compact submanifolds with boundary for $0 < i < k$ such that $\varphi(M_i) \subset \operatorname{Int} M_i$ and in $M_{i+1} - M_i$ the maximal invariant set is Λ_i. The existence of filtrations implies the following proposition, which clarifies why in the statement of stability results and conjectures we can either mention Axiom A (hyperbolicity of the nonwandering set and density of periodic orbits) or just hyperbolicity of the limit set or the chain recurrent set; see [N,1972;FS,1977].

Proposition. *If $L(\varphi)$ (or just $L^+(\varphi)$ or $L^-(\varphi)$) is hyperbolic and there are no-cycles on it, then $\Omega(\varphi) = L(\varphi)$ and so φ satisfies Axiom A (and there are no cycles on $\Omega(\varphi)$). If the chain recurrent set of φ is hyperbolic then φ satisfies Axiom A and the no-cycle condition.*

The existence of a filtration implies a global control on the nonwandering or limit set when we perturb the map: there appear no nonwandering points far from the original ones. This fact and the persistence of basic sets stated above imply the Ω-stability theorem [S,1970].

Theorem. *The set of diffeomorphisms satisfying Axiom A and the no-cycle property is open in Diff$^k(M)$ and its elements are Ω-stable.*

In the way of a converse to this theorem, we have

Theorem [P,1970]. *If φ is an Axiom A diffeomorphism and there is a cycle on $\Omega(\varphi)$, then φ is not Ω-stable. In fact, there are arbitrarily close diffeomorphisms $\tilde{\varphi}$ such that* $\text{Per}(\varphi) \not\subseteq \text{Per}(\tilde{\varphi})$. *A similar statement is true for diffeomorphism with hyperbolic limit sets.*

Corollary. *If φ and all nearby diffeomorphisms have their nonwandering sets (limit sets) hyperbolic, then they are Ω-stable. On the other hand, if φ or arbitrarily close diffeomorphisms exhibit homoclinic tangencies (see Chapter II or III) then φ is not Ω-stable.*

In view of the results above we formulate the following conjecture.

Ω-Stability Conjecture: A diffeomorphism φ is C^k Ω-stable if and only if it satisfies Axiom A and the no-cycle property.

Back to the stability conjecture, in the early seventies Robbin [R,1971] proved that diffeomorphisms satisfying Axiom A and the transversality condition are C^k stable for $k \geq 2$. Soon afterwards, de Melo [M,1973] proved the result for C^1 diffeomorphisms of surfaces using ideas close to [PS, 1970]. By 1976, Robinson completed the result for $k = 1$ in any dimension using an approach somewhat different from the previous ones. Before, in [R,1974], he had the corresponding version for flows with $k \geq 2$. In another paper [R,1973], he also proved that the transversality condition is necessary for C^k stability.

So, in both the stability and the Ω-stability conjectures it remained to show that hyperbolicity of the nonwandering set was necessary for either kind of stability. This was the missing fact to establish such a fundamental link between stability or Ω-stability and hyperbolicity of the nonwandering (or limit) set. From the begining it was clear that with the available knowledge this goal was probably beyond reach for C^k stability or Ω-stability when $k \geq 2$. In fact, it is still unknown that $\overline{\text{Per}(\varphi)} = \Omega(\varphi)$ if φ is C^2 Ω-stable;

for $k = 1$, this follows from Pugh's closing lemma [P,1967]. By 1980, Mañé concluded both conjectures for C^1 surface diffeomorphisms [M,1982]; independently, Liao [L,1980] and Sannami [S,1983] obtained this same result. For flows, Liao seems to have made substantial progress toward the same conclusion on 3-manifolds; in higher dimensions the question is still open. It is interesting to observe that the situation looks rather different for manifolds with boundary and flows leaving the boundary invariant: there exist singular horseshoes that are stable but not hyperbolic [LP,1986].

Finally, just about 20 years after it was proposed, Mañé in a remarkable paper [M,1988] presented a solution of the C^1 stability conjecture for diffeomorphisms in any dimension. His proof, however, did not include the C^1 Ω-stability conjecture because he needed the transversality condition (typical of stability but not of Ω-stability) to complete his arguments. This was done in [P,1988], arguing instead just with the no-cycle property.

Actually, one may ask if the following set of equivalences is true:

$$f \in \mathcal{D}^1(M) \iff f \text{ satisfies Axiom A and the no-cycle property} \iff f \text{ is } \Omega\text{-stable.}$$

Here, $\mathcal{D}^1(M)$ denotes the interior, with respect to the C^1 topology, of the set of f's whose periodic orbits are all hyperbolic. From the results above it only remained to show that if f is in $\mathcal{D}^1(M)$ then it satisfies Axiom A. The truth of this statement, and thus of the set of equivalences above, was recently and independently announced by Aoki [A,1990] and Hayashi [H,1990].

Closing this section, we want to pose two questions that are somewhat inspired by the results above and are relevant in the context of homoclinic bifurcations, the main topic of this text. It concerns differentiable arcs or one-parameter families φ_μ of C^k diffeomorphisms such that φ_μ satisfies Axiom A and the transversality condition for $\mu < \mu_0$ and φ_{μ_0} is Ω-unstable; such μ_0 is called the ***first bifurcation point*** of φ_μ.

The first question is: what types of bifurcation can occur for φ_{μ_0} if the family φ_μ is ***generic***, i.e. if the family φ_μ belongs to some ***residual (Baire second category)*** class of families ? We ***conjecture*** that in ***two dimensions***, we only have three possibilities: φ_{μ_0} has a nonhyperbolic periodic orbit, a homoclinic tangency, or a heteroclinic tangency involving periodic points in a cycle. ***In higher dimensions*** there are more cases, like ***homoclinic tangencies of basic sets:*** $W^s(x)$ tangent to $W^u(y)$ for x, y in the same basic set, neither point being necessarily periodic. A main difference is that the ***boundary*** of a basic set in two dimensions is made of stable and unstable manifolds of periodic orbits; see Appendix II.

The next question concerns a generic family φ_μ on a surface such that φ_μ is Morse-Smale for $\mu < \mu_0$ and there exist values $\tilde{\mu} > \mu_0$ arbitrarily near μ_0 such that $\varphi_{\tilde{\mu}}$ has infinitely many periodic orbits. The problem now is whether there is $\mu_1 > \mu_0$ near μ_0 so that φ_{μ_1} exhibits some homoclinic tangency associated to some periodic point (see Chapter III). In higher dimensions we formulate the question in terms of homoclinic bifurcations as in Chapter VII: one may persistently (open set of families) create homoclinic orbits without creating homoclinic tangencies at all!

We refer the reader to [NP,1976;NP,1973] where similar questions were studied and a version of the above conjecture posed.

§2. Sensitive-chaotic dynamics

The notion of "*chaos*" in dynamical systems, as opposed to theology and the usual meaning of total disorder, refers to a situation where (forward) orbits do not converge to a periodic or quasi-periodic orbit and where ***the evolution of the orbits has some degree of unpredictability*** or ***their behaviour is sensitive with respect to initial conditions***. Although this phenomenon was theoretically known, in particular for non-trivial hyperbolic attractors, it came to many as a surprise that it also appeared in numerically generated orbits of quite simple systems. Among the first such examples, which were investigated numerically, there were the Lorenz attractor [**L**, 1963], the logistic map [**CE**, 1980] and the Hénon map [**H**, 1976]—in fact all these systems depend on parameters and for a substantial set of parameter values these sensitive or chaotic phenomena appear. All these examples are non-hyperbolic and a big effort was needed to get some theoretical understanding of them. The last two examples play an important role in the dynamics at homoclinic bifurcations, see Chapters III, VI and VII. In this section we shall formalize the main notions involved in "chaos" and indicate instances where they occur in hyperbolic systems. We want to point

out that there are different formalizations of these notions (although there is agreement about the general flavour).

Here, we restrict ourselves to dynamical systems defined by a continuous map $\varphi\colon M \to M$, where we assume M to be a compact metric space—if M is not compact the discussion still makes sense if we restrict to points in M whose positive orbits have a compact closure and, hence, are bounded. We say that the (positive) ***orbit*** $\{x, \varphi(x), \varphi^2(x), \cdots\}$ of x is ***sensitive*** or ***chaotic*** if there is a positive constant $C > 0$ such that for any $q \in \omega(x)$ (for the definition see the previous section) and any $\varepsilon > 0$ there are integers $n_1, n_2, n > 0$ such that $d(\varphi^{n_1}(x), q) < \varepsilon$, $d(\varphi^{n_2}(x), q) < \varepsilon$, but $d(\varphi^{n_1+n}(x), \varphi^{n_2+n}(x)) > C$. We observe that an (asymptotically) (quasi) periodic orbit is not chaotic in the above sense: for such orbits, if $\varphi^{n_1}(x)$ and $\varphi^{n_2}(x)$ are close, then $\varphi^{n_1+n}(x)$ and $\varphi^{n_2+n}(x)$ remain close for all $n > 0$. Also, a sensitive orbit in the above sense is unpredictable to the extend that if we know that some y on the positive orbit of x is extremely close to $q \in \omega(x)$, this is not enough to predict, say within distance C, all future iterates of y. This last fact is closely related with the sensitive dependence on initial conditions discussed in the previous section. Similarly to the fact that in the stable set of a ***nontrivial hyperbolic attractor*** one has sensitive dependence on initial conditions, one can prove that most points in such a stable set have ***sensitive orbits***—the set of points with chaotic orbits even has ***total Lebesgue measure*** in the stable set; see [**ER**, 1985]. Not only in hyperbolic attractors there are sensitive orbits, but also in nontrivial hyperbolic sets of saddle type—this can easily be seen using symbolic dynamics as introduced for the horseshoe example in Chapter II. The difference is that in this last case, the set of points with sensitive orbits has Lebesgue measure zero.

For the case where M is a manifold, and hence the notion "Lebesgue measure zero" is defined, we say that the ***dynamical system*** defined by φ is ***sensitive***, or that φ has ***sensitive*** or ***chaotic dynamics***, if the set of points with sensitive or chaotic orbits has ***positive Lebesgue measure***.

Up to now, the notion of sensitive or chaotic dynamics has mainly been defined negatively (?!): namely in terms of unpredictability. From the numerical experiments, mentioned before, it was however apparent that certain aspects of the positive orbits were very predictable: for any initial point (in some open set), one always find numerically the *same* ω-limit set. Thus, chaos or sensibility as observed above should not be interpreted as ***total unpredictability***: instead, it should mean that most orbits of the system will finally be getting apart from each other and continue to wander around a "larger" but definite set. This leads to the definition of a ***strange attractor***. We say that a compact set $A \subset M$ is a ***strange attractor*** if there is an open set U with a subset $N \subset U$ of Lebesgue measure zero, such that for all $x \in U \setminus N$, $\omega(x) = A$ and the (positive) orbit of x is chaotic. (We allow for the exceptional set N because even for hyperbolic attractors a dense set of points, of Lebesgue measure zero, is attracted to periodic orbits; and ***as long as N has Lebesgue measure zero it should not be of influence on numerical experiments***). Another and even more common definition of ***strange attractor*** is to require A to have sensitive dependence on initial conditions on U; see the discussions in Chapter VII.

In many cases, including the numerical examples, one is not only interested in the persistence of phenomena under a (small) perturbation of the initial point (of a positive orbit) but also under a small perturbation of the map $\varphi: M \to M$. Intuitively, one says that the dynamics of φ is ***persistently sensitive or chaotic*** if small perturbations of φ have, with positive probability, sensitive dynamics. But, the problem with this intuitive notion is that there is no "natural" measure on the set of maps $\varphi: M \to M$. On the other hand, if we would require all small perturbations of φ to have sensitive dynamics (which is the case for a nontrivial hyperbolic attractor and whose dynamics is called for this reason ***fully persistently sensitive or chaotic***) we would exclude important cases like the ***logistic map*** for many values of the parameter. There is, however, one important instance in which this notion can be formally defined: if we are in a context where the notion of generic k-parameter unfoldings $\varphi_{\mu_1,\dots,\mu_k}$ of φ is defined (with $\varphi_{0,\dots,0} = \varphi$), we say that φ has persistently sensitive dynamics if for any such generic k-parameter unfolding, the $\mu = (\mu_1, \dots, \mu_k)$ values, for which φ_μ has sensitive dynamics, has positive Lebesgue measure.

In the last section of Chapter VI and in Chapter VII we shall discuss the consequences of our investigations of homoclinic bifurcations in terms of the above notions of sensitive or chaotic dynamics.

CHAPTER I

EXAMPLES OF HOMOCLINIC ORBITS IN DYNAMICAL SYSTEMS

We discuss a number of dynamical systems with (transverse) homoclinic orbits, just to have some examples in order to motivate the following chapters. First we need some definitions. We deal with diffeomorphisms $\varphi: M \to M$ of a compact manifold to itself. In this chapter it is enough to assume that φ is of class C^1, but for some of the later results we need φ to be C^2 or C^3. Also the compactness of M is not always needed—some of the examples in this chapter will be on $\mathbf{R}^2$.

We say that $p \in M$ is a hyperbolic fixed point of φ if $\varphi(p) = p$ and if $(d\varphi)_p$ has no eigenvalue of norm one. For such a hyperbolic fixed point, one defines the stable and the unstable manifold as

$$W^s(p) = \{x \in M | \varphi^i(x) \to p \text{ for } i \to +\infty\}$$

and

$$W^u(p) = \{x \in M | \varphi^i(x) \to p \text{ for } i \to -\infty\}.$$

According to the invariant manifold theorem (see [HPS, 1977] and also Appendix I) both $W^s(p)$ and $W^u(p)$ are injectively immersed submanifolds of M, are as differentiable as φ, and have dimensions equal to the number of eigenvalues of $(d\varphi)_p$ with norm smaller than one, respectively bigger than one. One can give the corresponding definitions for periodic points, i.e. fixed points of some power of φ.

If $\varphi: \mathbf{R}^n \to \mathbf{R}^n$ is a linear map with no eigenvalues of norm one, then the origin 0 is a hyperbolic fixed point and $W^s(0)$, $W^u(0)$ are complementary linear subspaces: $\mathbf{R}^n = W^s(0) \oplus W^u(0)$.

We say that if p is a hyperbolic fixed point of φ, q is ***homoclinic*** to p if $p \neq q \in W^s(p) \cap W^u(p)$, i.e. if $q \neq p$ and if $\lim_{i \to \pm\infty} \varphi^i(q) = p$ (this last form of the definition makes clear why Poincaré called such points "***bi-asymptotique***"). We say that q is a ***transverse homoclinic*** point if $W^s(p)$ and $W^u(p)$ intersect transversally at q, i.e. if

$$T_q(M) = T_q(W^s(p)) \oplus T_q(W^u(p)).$$

It is clear that ***linear diffeomorphisms*** have ***no homoclinic points.***

§1. Homoclinic orbits in a deformed linear map

We start with the linear map $\varphi: \mathbf{R}^2 \to \mathbf{R}^2$, $\varphi(x, y) = (2x, \frac{1}{2}y)$. The stable manifold is the y-axis, the unstable manifold is the x-axis. Next consider the composition $\Psi \circ \varphi$, where $\Psi: \mathbf{R}^2 \to \mathbf{R}^2$ is a diffeomorphism of the form

$$\Psi(x, y) = (x - f(x + y), y + f(x + y)),$$

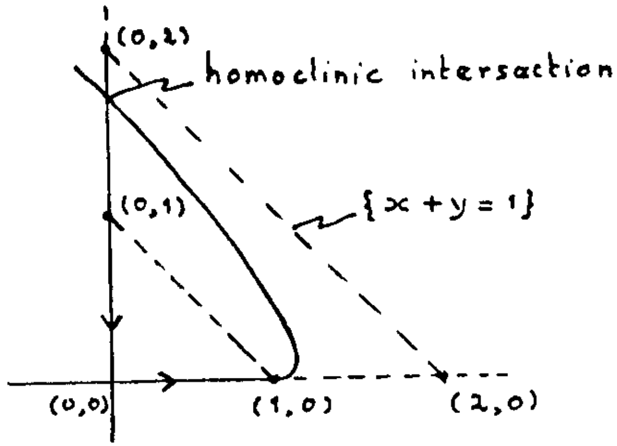
where f is some smooth function. This means that Ψ is pushing points along lines of the form $\{x + y = c\} = \ell_c$ over a distance which only dependes on c. We take f a smooth function which is zero on $(-\infty, 1]$ and such that $f(2) > 2$. In this case the stable and unstable manifold $W^s(0)$ and $W^u(0)$ for the diffeomorphism $\Psi \circ \varphi$ intersect outside the origin.

In fact, due to the construction, $\{(x, y) | x = 0, y \leq 2\}$ belongs to $W^s(0)$ and $\{(x, y) | x \leq 1, y = 0\}$ belongs to $W^u(0)$. Also

$$\Psi(\{(x, y) | 1 \leq x \leq 2, y = 0\})$$

belongs to $W^u(0)$. From this and the description of Ψ we obtain a homoclinic intersection, see the above figure. By choosing f appropriately we can produce a ***transverse*** homoclinic

point. Not much can be said about the global configuration of $W^s(0)$ and $W^u(0)$ but certainly this configuration will be very complicated, see later in this chapter.



§2. The pendulum

Our next example, the ***pendulum***, contains a line of non-transverse homoclinic points. Consider the differential equation

$$\ddot{\theta} = -\sin\theta, \quad \theta \in \mathbf{R}/2\pi,$$

which defines a system of ordinary differential equations on the annulus

$$\begin{cases} \dot{\theta} &= y \\ \dot{y} &= -\sin\theta \end{cases} \tag{1}$$

with $\theta \in \mathbf{R}/2\pi$, $y \in \mathbf{R}$.

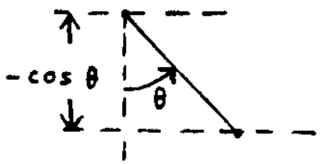
We take the time T map of this system, i.e. the diffeomorphism φ such that $\varphi(\theta, y) = (\tilde{\theta}, \tilde{y})$ whenever there is a solution $(\theta(t), y(t))$ of (1) with $(\theta(0), y(0)) = (\theta, y)$

and $(\theta(T), y(T)) = (\dot\theta\ \dot y)$. Then the fixed points of φ are $(\theta = 0, y = 0)$ and $(\theta = \pi, y = 0)$. The first is not hyperbolic (the eigenvalues of $(d\varphi)_{(0,0)}$ have norm one) but the second is: it has a one dimensional stable and a one dimensional unstable manifold. In order to determine the positions of these stable and unstable manifolds it is important to note that the function

$$E(\theta, y) = -\cos\theta + \frac{1}{2}y^2$$

is constant along solutions of (1): it is the energy, $-\cos\theta$ being the potential energy and $\frac{1}{2}y^2 = \frac{1}{2}\dot\theta^2$ being the kinetic energy. This means that both $W^u(\pi,0)$ and $W^s(\pi,0)$ are given by

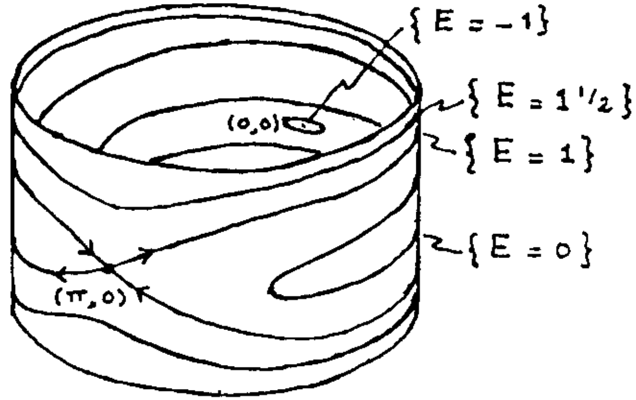
$$-\cos\theta + 1/2\ y^2 = 1$$



In the next figure this homoclinic line is indicated together with some other energy levels.

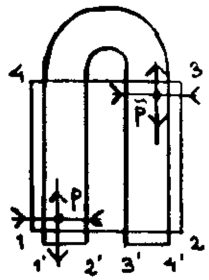
By a small perturbation of φ one can make $W^u(\pi,0)$ and $W^s(\pi,0)$ to intersect transversally (using a perturbation as in the first example, or refering to the proof of the Kupka-Smale theorem [S,1963;K,1964]). Such perturbations also arise as a consequence of external forcing, e.g. through replacing (1) by

$$\dot\theta = y$$
$$\dot y = -\sin\theta + \varepsilon \cdot \cos(2\pi t/T).$$

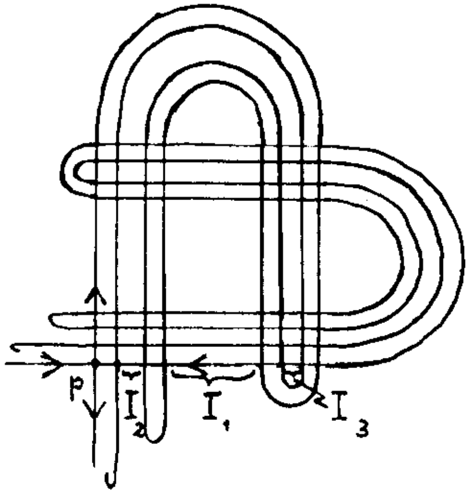


§3. The horseshoe

In the following example, the ***horseshoe***, see Smale [S,1965], we have transverse homoclinic points and still are fairly well able to describe globally the stable and the unstable manifold. In order to describe the diffeomorphism, let Q be a square in $\mathbf{R}^2$ and let φ map Q as indicated below, such that on both components of $Q \cap \varphi^{-1}(Q)$, φ is affine and preserves both horizontal and vertical directions, and such that 1,2,3 and 4 are mapped to 1', 2', 3' and 4'.



In Q, φ has two fixed points p and $\tilde{p}$ as indicated; we restrict our attention to p. Since φ is affine on $Q \cap \varphi^{-1}(Q)$, the stable and unstable manifolds $W^s(p)$ and $W^u(p)$, near p, are straight lines. In order to find the continuation one has to iterate φ^{-1} (for $W^s(p)$) and φ (for $W^u(p)$). Inside Q this gives pieces of straight lines, horizontal for $W^s(p)$ and vertical for $W^u(p)$. With a few iterations one gets the following picture.

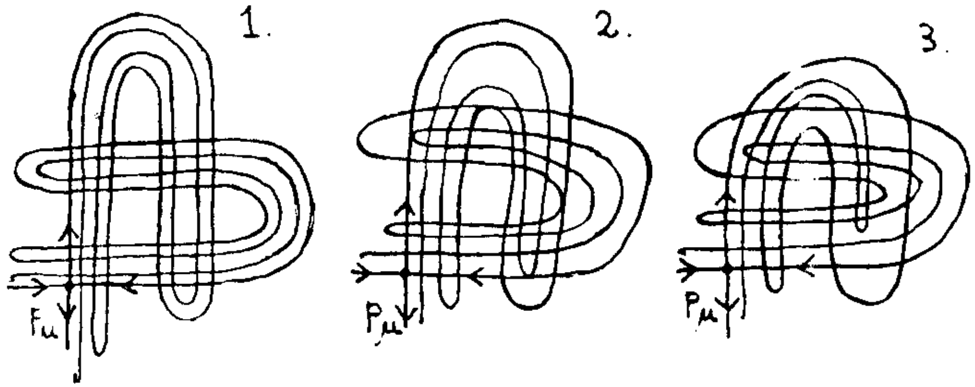


As a final remark on this example, observe that however far we iterate, the intervals $I_1, I_2, I_3 \subset W^s(p)$ will never be intersected by $W^u(p)$. There is in fact a countably infinite number of such open intervals, and $W^s(p) \cap W^u(p)$ consists of the boundary points of a Cantor set in $W^s(p)$ which is the complement of these intervals.

§4. A homoclinic bifurcation

We speak of a homoclinic bifurcation if in a one-parameter family of diffeomorphisms, a pair of homoclinic intersections collides, forms a tangency and then disappears, or, reversing the direction, if a pair of homoclinic points is generated after a tangency.

Such bifurcations are obtained from the previous example by composing φ with a map $(x, y) \mapsto (x, y - \mu)$, which slides the image, in particular the image of Q, down. In the next figures we show the effect of this sliding on the geometry of the stable and unstable manifold $W^s(p_\mu)$ and $W^u(p_\mu)$ for increasing values of μ. In the first one we have just the previous example.

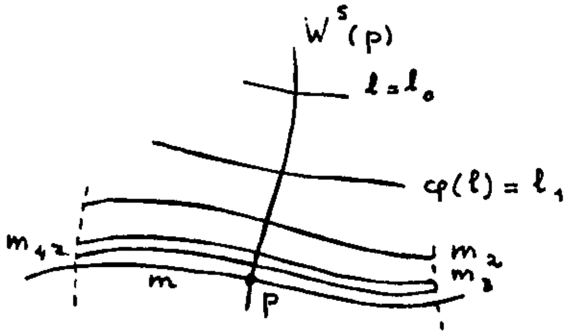


In the second figure one sees the first non-transverse homoclinic orbit (four iterations are indicated). From the third figure, it is clear that near one such homoclinic bifurcation there are many others—see also Chapter III.

§5. Concluding remarks

The complexity of the configuration of stable and unstable manifold in the examples is typical for the case where one has at least one transverse homoclinic point. Paraphrasing Poincaré, one can convince oneself of this complexity by trying to draw examples, keeping in mind that:

- $W^s(p)$ and $W^u(p)$ are φ-invariant, i.e. $\varphi(W^s(p)) = W^s(p)$ and $\varphi(W^u(p)) = W^u(p)$;
- $W^s(p)$ and $W^u(p)$ have no self intersections;
- near p, φ is well approximated by the linear map $(d\varphi)_p$, which leads to the following consequence (λ.lemma, [**P**,1969], see also Appendix 1): if ℓ is a smooth curve intersecting $W^s(p)$ transversally then the forward images $\ell_i = \varphi^i(\ell)$ contain compact arcs $m_i \subset \ell_i$ which approach differentiably a compact arc m in $W^u(p)$, as illustrated in the figure.



For higher or even infinite dimensional diffeomorphisms the situation is basically the same.

The first time that transverse homoclinic points were constructed was by Poincaré in his prize essay [**P**.1890]. The existence of these homoclinic points implied the non-convergence of certain power series expressions for solutions of a Hamiltonian system comparable with the Hamiltonian sytem describing the restricted 3-body problem. This indicated that certain qualitative information, like "stability", was unobtainable by these analytic power series methods.

Later it was realized that the dynamics of a diffeomorphism φ, or the topology of its orbits, shows a great complexity if and only if φ has some hyperbolic periodic point with a homoclinic intersection of its stable and unstable manifold. The same can be said about ***chaotic orbits***, see Chapter 0, Section 2: they occur if and only if there is some homoclinic

orbit (but a homoclinic orbit does not need to imply ***chaotic dynamics***). ***Although these last statements are in no way theorems (and can be expected to be true only in a generic sense), many facts discussed in these notes can actually be interpreted as partial results in this direction, see Chapter VII.*** On the other hand, we also discuss "stability results" concerning the dynamics in the presence of transverse homoclinic points. They deal with the not so infrequent situation that the dynamics of a diffeomorphism φ (with transverse homoclinic points), although very complicated, remains unchanged in a topological sense when φ is C^1 slightly perturbed, see Chapter 0.

CHAPTER II

DYNAMICAL CONSEQUENCES OF A TRANSVERSE HOMOCLINIC INTERSECTION

In this chapter we analyse the dynamical complexity due to one transverse homoclinic orbit. Although our discussion refers to the two-dimensional situation, the results and their proofs can be extended to arbitrary dimensions with minor modifications and even to Banach spaces (or Banach manifolds).

Since one of the main features of a transverse homoclinic intersection consists of the occurence of hyperbolic invariant sets, this chapter relies to some extent on Appendix 1, and even seems to overlap with the subject matter of that appendix. The emphasis however is different. Here we concentrate on the *geometric* properties of invariant sets near transverse homoclinic orbits which we show to be hyperbolic, and to do this we make use of some results from the *analytical* theory of hyperbolicity in Appendix I. For a more complete view on hyperbolic sets in dimension 2 one should also consult Appendix II on Markov partitions. To present the main ideas in a more transparent way, we will assume the diffeomorphisms in this chapter to be of class C^2 although the results are still true in the C^1 category.

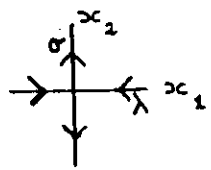
§1. Description of the situation—linearizing coordinates

Let $\varphi: M \to M$ be a C^2 diffeomorphism of a surface M to itself and let $p \in M$ be a hyperbolic fixed point of saddle type, i.e. $\varphi(p) = p$ and $(d\varphi)_p$ has two real eigenvalues λ

and σ with $0 < |\lambda| < 1 < |\sigma|$. For simplicity we assume that these eigenvalues are positive, so $0 < \lambda < 1 < \sigma$. From the theory of hyperbolicity (see Appendix I) we know that:

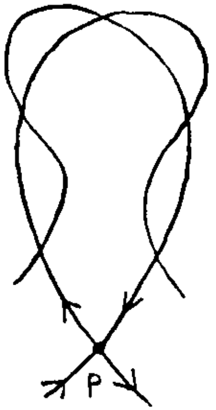
- the stable and unstable separatrices of p, $W^s(p)$ and $W^u(p)$, are C^2;
- there are C^1 linearizing coordinates in a neighbourhood U of p, i.e. C^1 coordinates x_1, x_2 such that $p = (0,0)$ and such that $\varphi(x_1, x_2) = (\lambda \cdot x_1, \sigma \cdot x_2)$.

This linearization follows at once from the existence of φ-invariant stable and unstable foliations near p which are of class C^1; see also [H,1964].

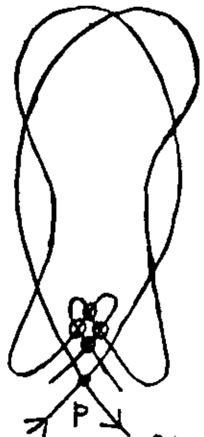


We assume that $W^s(p)$ and $W^u(p)$ have points of transverse intersection different from p—such points, or their orbits, are called ***homoclinic*** or ***bi-assymptotic*** to p. In the two-dimensional situation we consider mainly ***primary*** homoclinic points in order to simplify the figures and the geometric arguments. A homoclinic point q is primary if the arcs ℓ^u, joining p and q in W^u, and ℓ^s, joining p and q in W^s, form a double point free closed curve.

Note that whenever p has homoclinic points, it has primary homoclinic points—if all intersections of $W^u(p)$ and $W^s(p)$ are transverse, then the number of primary homoclinic orbits is finite. We also note that the notion of primary homoclinic orbit does not extend to dimensions greater than two. Still the results which we discuss below extend without much difficulty to the nonprimary or n-dimensional case.



all indicated homocli-
nic points are primary



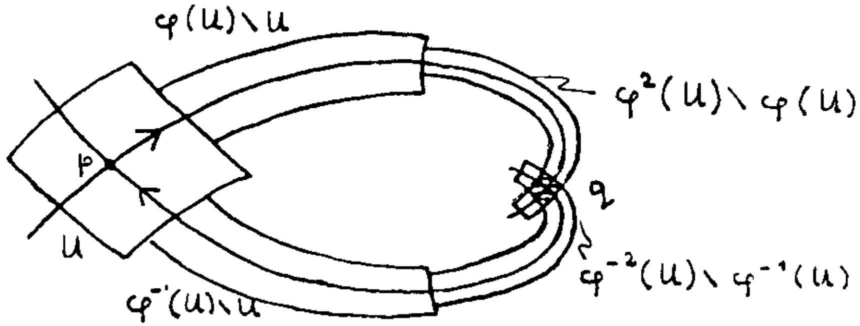
encircled
points not primary

Let the linearizing coordinates be defined on U and let their image be the square $(-1,+1)\times(-1,+1)\subset \mathbf{R}^2$. We consider extensions of the domain of definition of these linearizing coordinates. Identifying points in U with the corresponding points in $\mathbf{R}^2$, we have: if $\varphi^{-1}([\lambda,1)\times(-1,+1))\cap U=\phi$, we can extend the domain of the linearizing coordinates x_1,x_2 to $\varphi^{-1}([\lambda,1)\times(-1,+1))$ using the formulas

$$x_1=\lambda^{-1}\cdot(x_1\circ\varphi),\quad x_2=\sigma^{-1}\cdot(x_2\circ\varphi).$$

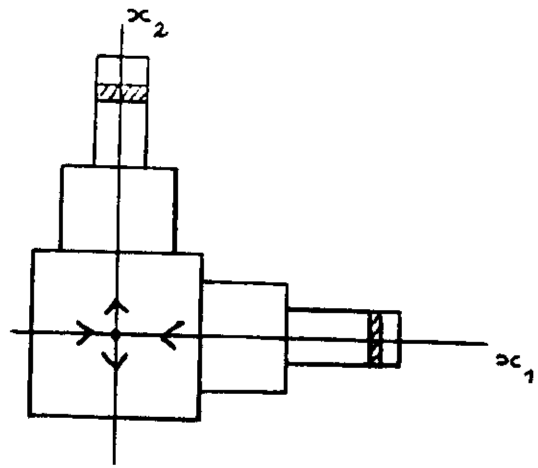
Repeating this construction one can extend the linearizing coordinates along any segment in $W^s(p)$ starting in p: one only has to take the original domain U sufficiently small. This follows from the fact that $W^s(p)$ has no self intersections. In the same way one can extend the domain of these linearizing coordinates along the unstable separatrix $W^u(p)$. Homoclinic intersections, however, form an obstruction to a simultaneous extension of such coordinates along both the stable and the unstable separatrix. In our situation where q is a primary homoclinic point, we extend the linearizing coordinates both along ℓ^u and ℓ^s, the arcs in $W^u(p)$ and $W^s(p)$ joining p and q; however, these coordinates will be bi-valued

near q as indicated below.



Situation in M, ≡ denotes the neighbourhood of q with bi-valued linearizing coordinates.

The figure below shows the situation in $\mathbf{R}^2$: the shaded area denotes the two sets corresponding to the above neighbourhood of q.



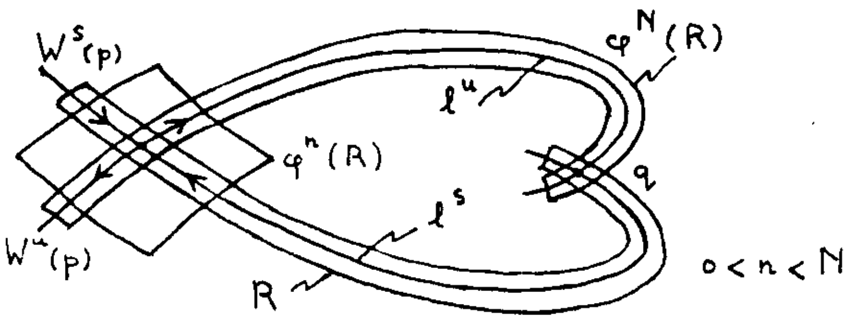
Now we consider in the domain of the extended coordinates a rectangle $R = \{-a \leq x_1 \leq b, -\alpha \leq x_2 \leq \beta\}$, $a, b, \alpha, \beta > 0$, containing ℓ^s, the arc in $W^s(p)$ joining p and q, and such that for some N:

- $R \cap \varphi^n(R)$ consists of one rectangle containing p for $0 \leq n < N$;
- $R \cap \varphi^N(R)$ consists of two connected components, one containning q, as indicated in the figure below, i.e.

$$\{x_1 = b, -\alpha \leq x_2 \leq \beta\} \cap \varphi^N(R) = \phi$$

$$\varphi^N(\{-a \leq x_1 \leq b, x_2 = \beta\}) \cap R = \phi.$$

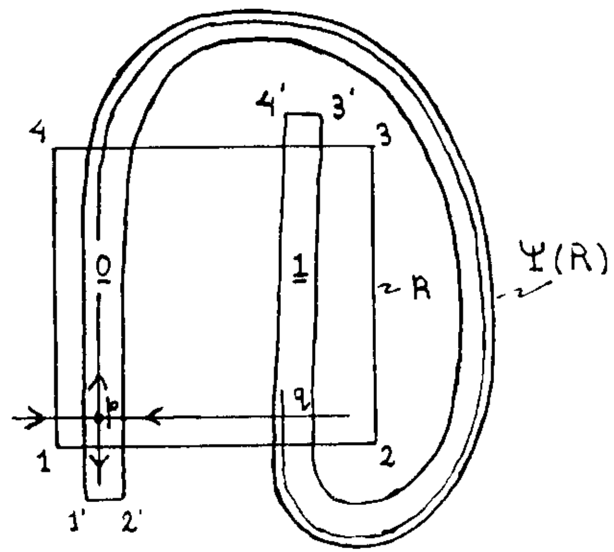
For what follows it is important that one can choose R so that N is arbitrarily big: just take β small. By taking β small and hence N big, $\varphi^N(R)$ will become a very narrow strip around ℓ^u. So, transversality of $W^u(p)$ and $W^s(p)$ at q implies transverse intersection of the sides of R and $\varphi^N(R)$.



The main object of interest in this chapter is the maximal invariant subset of R under φ^N, i.e. the set of those points $r \in R$ such that $\varphi^{k \cdot N}(r) \in R$ for all $k \in \mathbf{Z}$.

§2. The maximal invariant subset of R—topological analysis

From now on we denote φ^N by Ψ. We denote the *maximal invariant subset* of R under Ψ (satisfying the conditions in the previous Section) by $\Lambda = \{r \in R \mid \Psi^k(r) \in R$ for all $k \in \mathbf{Z}\}$. Denoting the corners of R by 1,2,3, and 4 and their images in $\Psi(R)$ by 1',2',3', and 4', the relative positions of R and $\Psi(R)$ are as indicated in the figure, i.e. the sides of R and $\Psi(R)$ intersect transversally and the topology of the positions of R its sides and its corners relative to their images under Ψ are as in the figure. We denote the components of $R \cap \Psi(R)$ by $\underline{0}$ and $\underline{1}$, $\underline{0}$ containing p and $\underline{1}$ containing q.



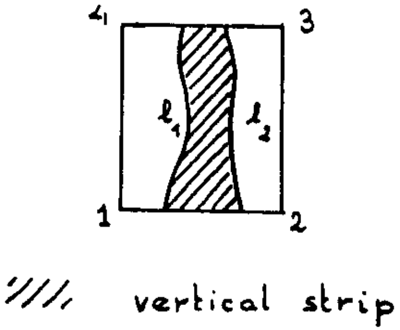
Theorem. *For any sequence $\{a_i\}_{i \in \mathbf{Z}}$, with $a_i = 0$ or 1, there is at least one point $r \in \Lambda$ such that $\Psi^i(r) \in \underline{a}_i$ for all $i \in \mathbf{Z}$.*

Proof: We call a closed subset $S \subset R$ a vertical strip if it is bounded (in R) by two disjoint continuous curves ℓ_1 and ℓ_2 connecting the side (1,2) with the side (3,4). If S is a vertical strip then $\Psi(S_j) \cap R$ contains two vertical strips, one in $\underline{0}$ and one in $\underline{1}$. Let now $\{a_i\}_{i\in\mathbf{Z}}$ be a sequence as in the theorem. We construct a nested sequence of vertical strips $S_0 \supset S_1 \supset S_2 \supset \cdots$: $S_0 = \underline{a}_0$; S_1 is the vertical strip $\Psi(\underline{a}_{-1}) \cap S_0$; S_2 is the vertical strip $\Psi^2(\underline{a}_{-2}) \cap S_1; \ldots ; S_\infty = \bigcap_{i\geq 0} S_i$.

For each point $r \in S_\infty$, $\Psi^{-i}(r) \in \underline{a}_{-i}$, $i \geq 0$.

Horizontal strips are similarly defined and we have horizontal strips $T_1 \supset T_2 \supset T_3 \supset \cdots$ such that for $r \in T_\infty = \bigcap_{i\geq 1} T_i$, $\Psi^i(r) \in \underline{a}_i$ for all $i \geq 1$. Now $S_\infty \cap T_\infty \neq \phi$. Otherwise, for some $i_0, S_{i_0} \cap T_{i_0} = \phi$, but S_{i_0} contains a line from side (1,2) to the side (3,4) and T_{i_0} contains a line from (1,4) to (2,3). These lines have to intersect.

For any point $r \in S_\infty \cap T_\infty$, $\Psi^i(r) \in \underline{a}_i$ for all $i \in \mathbf{Z}$. From this, it also follows that $r \in \Lambda$. ∎



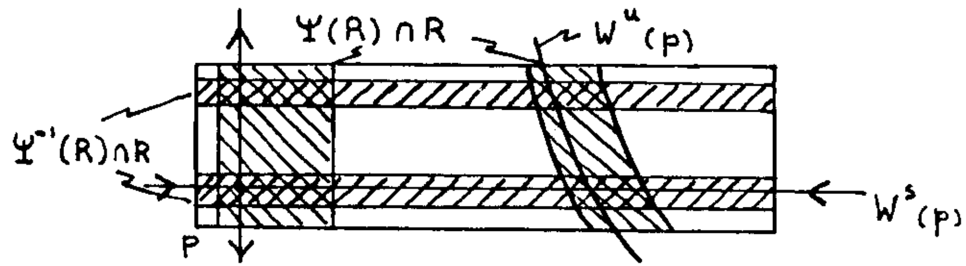
§3. The maximal invariant subset of R–hyperbolicity and invariant foliations

In this section we impose more conditions on $\Psi = \varphi^N$ restricted to R. In the linearizing coordinates on a neighbourhood of ℓ^s, the arc in $W^s(p)$ joining p and q, we

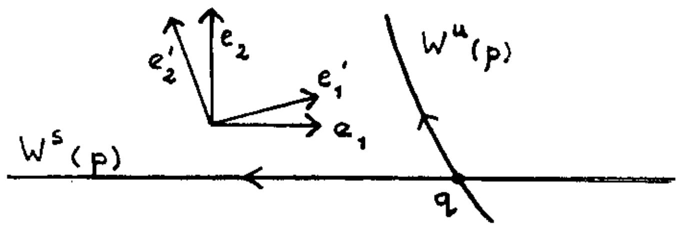
have

$$R = \{-a \leq x_1 \leq b, -\alpha \leq x_2 \leq \beta\},$$

see Section 1. We only have to describe Ψ in those points of R which are mapped back into R, i.e. in $\Psi^{-1}(R) \cap R$. In the component of $\Psi^{-1}(R) \cap R$ containing p and q, Ψ is linear and in fact $\Psi(x_1, x_2) = (\lambda^N x_1, \sigma^N x_2)$ with $0 < \lambda < 1 < \sigma$.



The other component of $\Psi^{-1}(R) \cap R$ is mapped to the component of $R \cap \Psi(R)$ containing q. This component of $R \cap \Psi(R)$ is the region where the linearizing coordinates, constructed in Section 1, were bi-valued, or rather where we have apart from the linearizing coordinates following $W^s(p)$, and which are in the above figure the Cartesian coordinates of the plane, also the linearizing coordinates following $W^u(p)$. We denote by e_1, e_2 the coordinate vector fields of the linearizing coordinates following $W^s(p)$ and by e_1', e_2' the coordinate vector fields of the linearizing coordinates following $W^u(p)$; see the figure below.



For r in the component of $R \cap \Psi^{-1}(R)$ which is mapped on a neighbourhood of q, we have $(d\Psi)e_1(r) = \lambda^N \cdot e_1'(\Psi(r))$ and $(d\Psi)e_2(r) = \sigma^N \cdot e_2'(\Psi(r))$.

Due to the transversality of $W^u(p)$ and $W^s(p)$ and to the thinness of $\Psi(R)$, for N big, e_1 and e_2' are linearly independent. Also, by choosing R and $\Psi(R)$ thin, we may assume that the matrix transforming e_1, e_2 into e_1', e_2' (or its inverse) is almost constant.

Theorem. *For R sufficiently thin, and hence N big, the maximal invariant subset $\Lambda = \bigcap_{n\in\mathbf{Z}} \Psi^n(R)$ in R is hyperbolic. (The technical definition of hyperbolicity is given in Appendix I).*

Proof: A continuous cone field C on $R \cap \Psi(R)$ is a map which assigns to each $r \in R\cap\Psi(R)$ a two-sided cone $C(r)$ in $T_r(M)$, given by two linearly independent vectors $w_1(r), w_2(r)$:

$$C(r) = \{v \in T_r(M) | v = a_1 \cdot w_1(r) + a_2 \cdot w_2(r) \text{ with } a_1 \cdot a_2 \geq 0\}.$$

Continuity of C means that w_1 and w_2 depend continuously on r. Let O_r denotes the zero vector in $T_r(M)$. An unstable cone field is a continuous cone field on $R \cap \Psi(R)$ such that

- for each $r \in R \cap \Psi(R) \cap \Psi^{-1}(R)$,
  $\overline{(d\Psi)(C(r))} \subset \operatorname{Int}(C(\Psi(r))) \cup \{O_r\}$,
- for each $r \in R \cap \Psi(R) \cap \Psi^{-1}(R)$ and $v \in C(r)$,
  $\|d\Psi(v)\| \geq \nu \|v\|$, for some $\nu > 1$, where both norms are taken with respect to the basis e_1, e_2.

Below we construct such an unstable cone field. From the existence of such a cone field it follows that there is a continuous direction field $E^u(r)$, defined for $r \in \bigcap_{i\geq 0} \Psi^i(R)$, such that

- $E^u(r) \subset C(r)$;
- $d\Psi$ maps $E^u(r)$ to $E^u(\Psi(r))$;
- for some $\nu > 1$, and all $v \in E^u(r)$ with $r \in \bigcap_{i\geq -1} \Psi^i(R)$, $\|d\Psi(v)\| \geq \nu \cdot \|v\|$.

E^u, restricted to Λ, is obtained by taking the intersections of the forward images of the cone field C under $d\Psi$.

Replacing Ψ by Ψ^{-1}, one constructs in the same way a stable cone field and the direction field E^s, which is invariant under and contracted by $d\Psi$. Then $T_\Lambda(M) = E^u_\Lambda \oplus e^s_\Lambda$ is the required splitting for hyperbolicity—see Appendix I.

Now we come to the construction of the unstable cone field on $R \cap \Psi(R)$. In the component of $R\cap\Psi(R)$ containing p we simply take cones around e_2 extending 45° to both sides. In the other component of $R\cap\Psi(R)$ there is (assuming R and $\Psi(R)$ sufficiently thin) an angle α, smaller than 90°, so that for each point r in that component of $R \cap \Psi(R)$, the cone around $e_2(r)$, extending over an angle α to both sides, contains $e_2'(r)$ in its interior. This is due to the fact that $e_1(r)$ and $e_2'(r)$ are linerarly independent. The unstable cone field C is just defined as the field of cones, centred on e_2 and extending 45°, respectively α, to both sides of e_2 depending on the component of $R \cap \Psi(R)$.

In order to show that this cone field has the required properties, we introduce constants A, B and B' so that: whenever $r \in R \cap \Psi(R)$ and $v = v_1 \cdot e_1(r) + v_2 \cdot e_2(r) \in T_r(M)$ then for $v \in C(r)$ we have $|v_1| \leq A \cdot |v_2|$, on the other hand whenever $|v_1| \leq B \cdot |v_2|$, then $v \in C(r)$; whenever $v = v_1' \cdot e_1'(r) + v_2' \cdot e_2'(r)$ and $|v_1'| \leq B' \cdot |v_2'|$, then $v \in C(r)$. If N is so big that $\left|\frac{\lambda}{\sigma}\right|^N \cdot A < \min(B, B')$, then $d\Psi$ maps cones in the interior of cones. Also for N sufficiently big, the lengths of vectors in our cones are strictly increased by $d\Psi$.

So for N big enough our cone field has the required properties. *But* the cone field was constructed *after* choosing N, Ψ is defined in terms of N: $\Psi = \varphi^N$, and the domain of the cone field is defined in terms of Ψ: domain $= R \cap \Psi(R)$. *However*, the way to raise N is to make R thinner. This decreases the domain where the cone field has to be defined. The fact that Ψ changes from φ^N to $\varphi^{N'}$, $N' > N$, has no influence on the arguments: the vector fields e_1, e_2, e_1' and e_2' do not change. So R and N may be ajusted afterwards. This completes the proof of the hyperbolicity of Λ. ∎

Observe that we proved slightly more: there are vector fields $e^u \in e^u_\Lambda$ and $e^s \in e^s_\Lambda$ and a constant $\nu > 1$, such that for all $r \in \Lambda$, $\|d\Psi(e^u(r))\| \geq \nu \cdot \|e^u(r)\|$ and $\|d\Psi(e^s(r))\| \leq \nu^{-1} \cdot \|e^s(r)\|$.

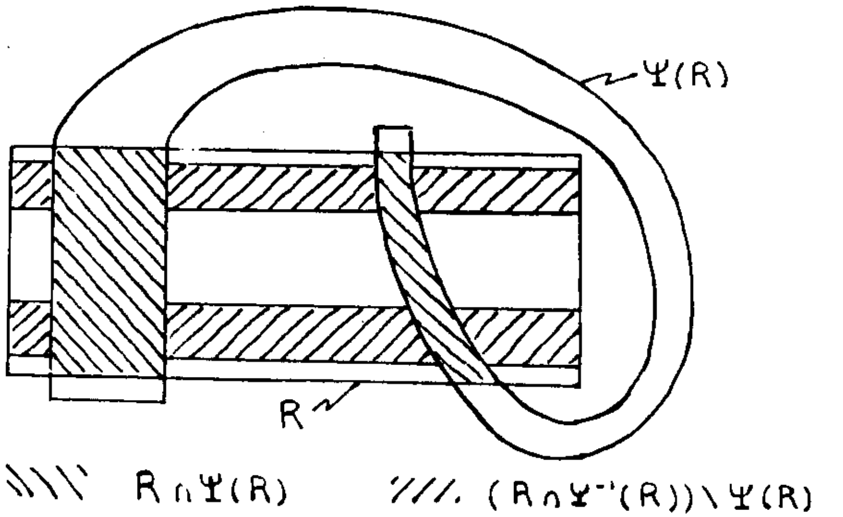
Now we come to the second subject of this section. The cone fields just constructed will now be used to construct the stable and the unstable foliation. We only describe the construction of the unstable foliation. First we present the definition.

An *unstable foliation* for $\Lambda = \bigcap_{i \in \mathbb{Z}} \Psi^i(R)$ is a foliation $\mathcal{F}^u$ of a *neighbourhood* of Λ (here we take $R \cap \Psi(R))$ such that

1. for each $r \in \Lambda$, $\mathcal{F}^u(r)$, the leaf of $\mathcal{F}^u$ containing r, is tangent to $E^u(r)$;
2. for each r, sufficiently near Λ, $\Psi(\mathcal{F}^u(r)) \supset \mathcal{F}^u(\Psi(r))$.

We require the tangent directions of leaves of $\mathcal{F}^u$ to vary continuously.

Construction of the unstable foliation: We recall the relative positions of $R, \Psi(R)$, and $\Psi^{-1}(R)$ in the figure below.



We take a C^2 foliation $\tilde{\mathcal{F}}^u$ (not yet the unstable foliation) on $(\Psi(R) \cup \Psi^{-1}(R)) \cap R$ so that:

1. in $R \cap \Psi(R)$ the tangent directions of leafs are contained in the unstable cones;

2. the image under Ψ of leaves in $(R \cap \Psi^{-1}(R)) \setminus \Psi(R)$ have tangent directions contained in the unstable cones;
3. the four arcs of $\partial R \cap \Psi^{-1}(R)$ are leafs of $\tilde{\mathcal{F}}^u$, the union of these four arcs denoted by e_0;
4. the four arcs of $\partial(\Psi(R)) \cap R$ are leafs of $\tilde{\mathcal{F}}^u$, the union of these four arcs is denoted by e_1;
5. Ψ maps leaves of $\tilde{\mathcal{F}}^u$ near e_0 to leaves of $\tilde{\mathcal{F}}^u$ near e_1.

Since all the cones of the unstable cone field are centered around the vertical vector field e_2 and contain, (where defined) e_2', it is clear that such a foliation $\tilde{\mathcal{F}}^u$ exists.

For foliations as described above we define an operator Ψ_* as follows:
in $(R \cap \Psi^{-1}(R)) \setminus \Psi(R)$, the leaves of $\tilde{\mathcal{F}}^u$ and $\Psi_*(\tilde{\mathcal{F}}^u)$ are the same;
in $(R \cap \Psi(R))$, the leaves of $\Psi_*(\tilde{\mathcal{F}}^u)$ are connected components of Ψ-images of leaves of $\tilde{\mathcal{F}}^u$ intersected with $(R \cap \Psi(R))$.

Due to the above conditions 3, 4 and 5, $\Psi_*(\tilde{\mathcal{F}}^u)$ is also C^2. From invariant manifold theory it follows that the limit

$$\lim_{i \to \infty} \Psi_*^i(\tilde{\mathcal{F}}^u) = \mathcal{F}^u$$

exists. This limit depends on the choice of the "initial foliation" $\tilde{\mathcal{F}}^u$. The limit is C^1; if however φ is C^3 then this limit is $C^{1+\epsilon}$, see Appendix I.

Observe that we can extend our vector fields e^u and e^s in E^u, respectively E^s, to tangent vector fields of $\mathcal{F}^u$ and $\mathcal{F}^s$ ($\mathcal{F}^s$ is just an unstable foliation for Ψ^{-1}) so that for some constant $\tilde{\nu} > 1$, and all $r \in R \cap \Psi(R) \cap \Psi^{-1}(R)$,

$$\|d\Psi(e^u(r))\| \geq \tilde{\nu} \cdot \|e^u(r)\|$$

and

$$\|d\Psi(e^s(r))\| \leq \tilde{\nu}^{-1} \cdot \|e^s(r)\| .$$

Stable and unstable foliations can be constructed for any basic set of a C^1 diffeomorphism in dimension 2 [**M**,1973]. In higher dimensions the existence of such foliations constitute an interesting open problem except when the basic set is zero dimensional [**P**,1983];

for futher details see Appendix I. We stress that in this comment we are insisting on foliations defined in a neighborhood of the basic set! Sometimes in the literature one also refers to ***stable*** (or ***unstable***) ***foliation*** as the set of stable manifolds ("leaves") through the points of the basic set, and in this sense the foliation always exist for any hyperbolic set; see Appendix I.

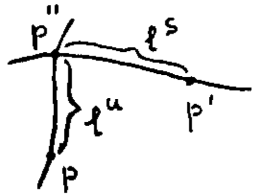
§4. The maximal invariant subset of R–structure

We divide Λ, the maximal invariant subset of R, in ***blocks***. For each sequence $A = (a_{-k}, a_{-k+1}, \cdots, a_{k-1}, a_k)$ with $a_i = 0$ or 1, we define the A-block as $\Lambda_A = \{r \in \Lambda | \Psi^i(r) \in \underline{a}_i$ for $i = -k, \cdots, k\}$; we call k the ***radius*** of A. The ***diameter*** of Λ_A is just the supremum of the distances between points in Λ_A, measured in $\mathbf{R}^2$.

As we saw in the last section, expansions and contractions of vectors along unstable, respectively stable, foliations are at least by a factor $\tilde{\nu} > 1$, respectively $\tilde{\nu}^{-1}$. Let c be the maximal length of a component of a stable or unstable leaf in $R \cap \Psi(R)$.

Proposition. *Let Λ_A be an A-block and let A have radius k. Then the diameter of Λ_A is at most $2 \cdot c \cdot \tilde{\nu}^{-k}$ (for the definition of c, $\tilde{\nu}$ see above).*

Proof: For any two points p, p' in the same component of $R \cap \Psi(R)$ there are unique arcs $\ell^u(p, p'')$ and $\ell^s(p', p'')$ in leaves of $\mathcal{F}^u$, respectively $\mathcal{F}^s$, joining p, respectively p', and the intersection p'' of the unstable leaf through p and the stable leaf through p'. When p, p' are both in Λ_A, then this whole configuration will remain in the same component of $R \cap \Psi(R)$ when we apply Ψ^i, $i = -k, \cdots, +k$. This implies that the lengths of $\ell^u(p, p'')$ and $\ell^s(p', p'')$ are both at most $c \cdot \tilde{\nu}^{-k}$. This implies the proposition. ∎



From the above proposition and the result of Section 2, we obtain:

Theorem. *The size of an A-block Λ_A goes to zero as the radius of A goes to infinity. For each infinite sequence $\cdots, a_{-2}, a_{-1}, a_0, a_1, a_2, \cdots, a_i = 0$ or 1, there is exactly one point $r \in \Lambda$ such that $\Psi^i(r) \in \underline{a}_i$ for all i. There is a homeomorphism $h: \Lambda \to (\mathbf{Z}_2)^{\mathbf{Z}}$ (product topology on $(\mathbf{Z}_2)^{\mathbf{Z}}$) such that for $r \in \Lambda$, $h(r) = \cdots, a_{-2}, a_{-1}, a_0, a_1, a_2, \cdots$ with $\Psi^i(r) \in \underline{a}_i$. If $\sigma: (\mathbf{Z}_2)^{\mathbf{Z}} \to (\mathbf{Z}_2)^{\mathbf{Z}}$ is the shift operator, i.e. $\sigma(\{a_i\}_{i\in\mathbf{Z}}) = \{a'_i\}_{i\in\mathbf{Z}}$ with $a'_i = a_{i+1}$, then*

$$\begin{array}{ccc} \Lambda & \xrightarrow{\Psi|\Lambda} & \Lambda \\ \big\downarrow h & & \big\downarrow h \\ (\mathbf{Z}_2)^{\mathbf{Z}} & \xrightarrow{\sigma} & (\mathbf{Z}_2)^{\mathbf{Z}} \end{array}$$

commutes.

Remark 1: It follows from the above theorem and its proof that if $\tilde{\Psi}$ is C^1 close to Ψ, then the same conclusions hold for the maximal invariant set $\tilde{\Lambda}$ of $\tilde{\Psi}$ in R. Namely, if C is an unstable cone field for Ψ whose domain is slightly extended beyond $R \cap \Psi(R)$, we conclude that C is also an unstable cone field for $\tilde{\Psi}$ if $\tilde{\Psi}$ is sufficiently C^1 close to Ψ. Then all the above arguments apply, with the obvious modifications, to $\tilde{\Psi}$. This implies that for such $\tilde{\Psi}$ and $\tilde{\Lambda}$, there is a conjugacy $H: \Lambda \to \tilde{\Lambda}$, i.e. a homeomorphism such that the diagram below commutes.

$$\begin{array}{ccc} \Lambda & \xrightarrow{\Psi|\Lambda} & \Lambda \\ H\big\downarrow & & \big\downarrow H \\ \bar{\Lambda} & \xrightarrow{\bar{\Psi}|\bar{\Lambda}} & \bar{\Lambda} \end{array}$$

Remark 2: The periodic points are dense in Λ. This follows from the corresponding statement for $(\mathbf{Z}_2)^{\mathbf{Z}}$ and σ: any finite sequence can be completed to an infinite periodic sequence. It is clear that all these periodic points are of saddle type (one expanding and one contracting direction). We add one more observation about these periodic orbits to be used in the next chapters. In general we say that a fixed point p of a diffeomorphism φ is *dissipative* if $|\det(d\varphi)(p)| < 1$. The same applies to periodic points, say of period k: just replace φ by φ^k. Now if the fixed point p with which we started this Chapter (see Section 1) is dissipative, then, for R sufficiently thin (or N big), all the periodic points in Λ will be dissipative. If $p' \in \Lambda$ is a periodic point of Ψ, and hence of φ, and if R is thin then p' has most points of its orbit (under iteration of φ) in a small neighbourhood of p. The dissipativeness then follows from

$$\det(d\varphi^k)_{p'} = \prod_{i=0}^{k-1} \det(d\varphi)_{\varphi^i(p')}.$$

Remark 3: It also follows from the above constructions that Λ has "local product structure" in the sense that if $r, r' \in \Lambda$ are in the same component of $R \cap \Psi(R)$, then the intersection of the stable leaf $\mathcal{F}^s(r)$ through r and the unstable leaf $\mathcal{F}^u(r')$ through r' also belongs to Λ. In fact, if $h(r) = \{a_i\}_{i\in\mathbf{Z}}$ and $h(r') = \{a'_i\}_{i\in\mathbf{Z}}$, then $\mathcal{F}^s(r) \cap \Lambda$ corresponds to the sequences

$$\{\{b_i\}_{i\in\mathbf{Z}} | b_i = a_i \text{ for } i \geq 0\}$$

and $\mathcal{F}^u(r') \cap \Lambda$ corresponds to the sequences

$$\{\{b'_i\}_{i\in\mathbf{Z}} | b'_i = a'_i \text{ for } i \leq 0\}$$

Since r, r' are in the same component of $R \cap \Psi(R)$, $a_0 = a'_0$, so that the point in $\mathcal{F}^s(r) \cap \mathcal{F}^u(r')$ corresponds to the sequence

$$\cdots, a'_{-2}, a'_{-1}, a'_0 = a_0, a_1, a_2, \cdots$$

Remark 4: A point $r \in \Lambda$ has a *chaotic orbit* if its symbolic sequence $h(r) = \{a_i\}$ is not eventually periodic, i.e., for no n, $a_n, a_{n+1}, \ldots$ is periodic. So in Λ, orbits are either asymptotically periodic or chaotic.

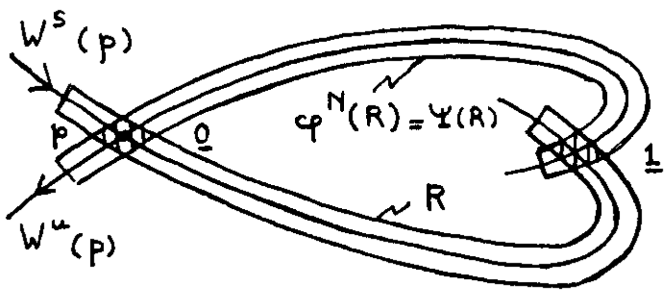
§5. Conclusions for the dynamics near a transverse homoclinic orbit

We return to the diffeomorphism φ (see Section 1) and discuss the consequences of the results we have obtained in Sections 2 to 4 for $\Psi = \varphi^N$. We have analysed the maximal invariant subset Λ in R under the map Ψ; what about the map φ?

This set Λ is contained in $R \cap \Psi(R)$ whose components are denoted by $\underline{0}$ and $\underline{1}$. A corresponding invariant set for φ is defined as $\hat{\Lambda} = \bigcup_{i=0}^{N-1} \varphi^i(\Lambda)$.

Proposition. *The set $\hat{\Lambda}$, as defined above is the disjoint union of $\{p\}$, $\Lambda - \{p\}$, $\varphi(\Lambda - \{p\}), \cdots, \varphi^{N-1}(\Lambda - \{p\})$.*

Proof: We only have to show that for $0 < i < N$, $\Lambda \cap \varphi^{-i}(\Lambda) = \{p\}$. In fact let $r \in \Lambda$, and $\varphi^i(r) \in \Lambda$ for some $0 < i < N$. Then $r \notin \underline{1}$ and also $\varphi^{k\cdot N}(r) = \Psi^k(r)$ has the same properties, i.e $\Psi^k(r) \in \Lambda$ and $\varphi^i(\Psi^k(r)) \in \Lambda$. This implies that $\Psi^k(r) \in \underline{0}$ for all k and hence that $r = p$. This proves the proposition. ∎



It is clear that $\hat{\Lambda}$ is a hyperbolic set for φ, and that the periodic orbits are dense in $\hat{\Lambda}$. As we have observed before, a (transverse) homoclinic orbit implies great complexity of the patterns formed by the corresponding separatrices. In this direction we can prove.

Proposition. *In the above situation, $\hat{\Lambda}$ is contained in the closure of both the stable and the unstable manifold of p.*

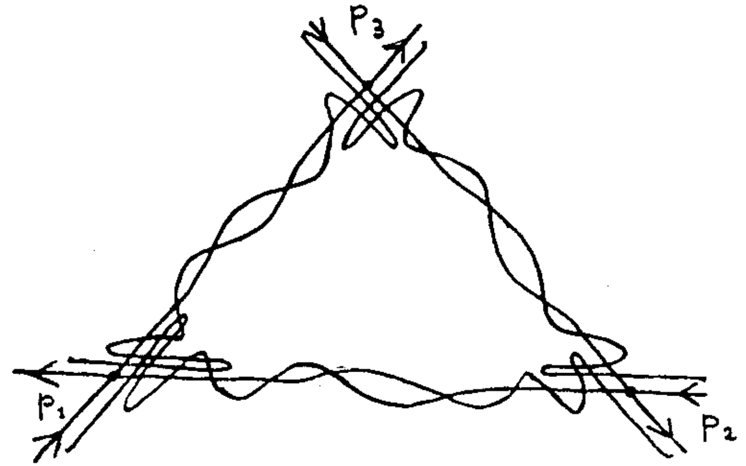
Proof: Since the periodic orbits are dense in $\hat{\Lambda}$ it is enough to prove that each periodic point of $\hat{\Lambda}$ is contained in $\overline{W^s(p)}$ (and in $\overline{W^u(p)}$). Since $\overline{W^s(p)}$ is invariant under φ, it is enough to prove that the periodic points of Λ are in $\overline{W^s(p)}$. For any periodic point $r \in \Lambda$, the unstable separatrix $W^u(r)$ contains the leaf of the unstable foliation through r and hence intersects $W^s(p)$. Then it follows (iterate φ^{-1}) that this periodic point r is contained in the closure of $W^s(p)$. In the same way one proves that it is contained in the closure of $W^u(p)$. ∎

§6. Homoclinic points of periodic orbits

Let again $\varphi: M \to M$ be a diffeomorphism but now with a periodic orbit of period k: $\{p_0, p_1, \cdots, p_{k-1}\}, \varphi(p_i) = p_j$ where $j = i + 1 \pmod{k}$. We assume that this periodic orbit is of saddle type. Stable and unstable manifolds are denoted by $W^s(p_i)$ and $W^u(p_i)$. There are two types of homoclinic orbits, namely intersections of $W^s(p_i)$ and $W^u(p_i)$—they are just homoclinic orbits of a hyperbolic saddle fixed point for φ^k—and intersections of $W^s(p_i)$ and $W^u(p_j)$, $i \neq j$. For $t = j - i$, we then have also intersections of $W^s(p_j)$ and $W^u(p_{j+t})$ etc... This means that we get something like a cycle whose "period" is the smallest number ℓ such that $\ell \cdot t$ is a multiple of k. If the intersection of $W^s(p_i)$ and $W^u(p_j)$ is transverse, so are the intersection of $W^s(p_j)$ and $W^u(p_{j+t})$, of $W^s(p_{j+t})$ and $W^u(p_{j+2t})$, etc... By the λ-lemma [**P**,1969], see also Appendix 1, this means that $W^s(p_i)$ is accumulating on $W^s(p_j)$ and hence intersecting $W^u(p_{j+t})$ transversally, hence

accumulating on $W^s(p_{j+t})$ etc., so that we finally get a transverse intersection of $W^s(p_i)$ with $W^u(p_i)$ anyway.

An example of this last phenomenon occurs in any generic 2-parameter family of diffeomorphisms $\varphi_\mu: \mathbf{R}^2 \to \mathbf{R}^2$, $\mu \in \mathbf{R}^2$, such that $\varphi_0(0) = 0$ and such that $(d\varphi_0)_0$ has eigenvalues $e^{\pm 2\pi i/3}$—this is the subharmonic bifurcation with resonance 1:3 [**A**,1980]. Stable and unstable separatrices are then as indicated below.



§7. Transverse homoclinic intersections in arbitrary dimensions

As we remarked in the beginning of this chapter, the results and proofs can be extended to diffeomorphisms $\varphi: M \to M$, where M is an n-dimensional manifold. So one obtains:

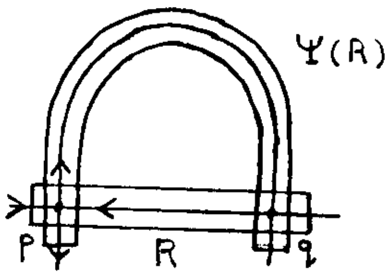
Theorem. *Let $\varphi: M \to M$ be a C^1 diffeomorphism with a hyperbolic fixed point p. Let q be a point of transverse intersection of $W^u(p)$ and $W^s(p)$. Then there is a neighbourhood*

U of the closure of the orbit $\overline{O(q)} = \overline{\bigcup_{i \in \mathbf{Z}} \varphi^i(q)}$ *such that the maximal invariant set* $\hat{\Lambda}$ *under* φ *in* U *is a nontrivial hyperbolic set (see Appendix I). Also, there are neighbourhoods* V_p *and* V_q *of* p *and* q *and there is an integer* N, *such that the maximal invariant set* Λ *under* φ^N *in* $V = V_p \cup V_q$ *is also a nontrivial hyperbolic set and such that* $\varphi^N|\Lambda$ *is conjugated with the shift on* $(\mathbf{Z}_2)^{\mathbf{Z}}$ *as in Section 4.*

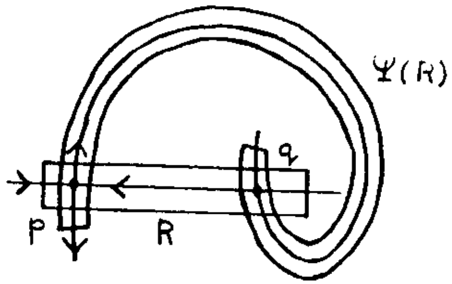
§8. Historical note

The main ideas in this chapter were developed by Poincaré [P,1890], who realized that homoclinic points are accumulated by other homoclinic points, by G.D. Birkhoff [B,1935] who showed that homoclinic points are accumulated by periodic points, and by S. Smale [S,1965] who essentially obtained the main theorem of Section 4.

The maximal invariant set in R is often called a horseshoe and the map $\Psi|R$ a horseshoe map. Due to the topology of R^2 there are two types of transverse homoclinic orbits but for both cases the analysis is the same; our figures refer to the less conventional case in which one does not "see" a horseshoe. In the conventional case one has



instead of



CHAPTER III

HOMOCLINIC TANGENCIES: CASCADE OF BIFURCATIONS, SCALING AND QUADRATIC MAPS

In this chapter we begin to discuss the unfolding of homoclinic tangencies for one-parameter families $\{\varphi_\mu\}$ of diffeomorphisms of a manifold M. As before, for simplicity of presentation, we shall consider here M to be a two-dimensional surface. In the last two sections we shall also assume the diffeomorphism to be ***locally dissipative***: the product of the eigenvalues at the saddle associated to the homoclinic tangency is smaller than one. The corresponding results in higher dimensions are true if we require the diffeomorphism to have ***only one expanding eigenvalue*** at the saddle associated to the homoclinic tangency and the ***product of any two eigenvalues to have norm smaller than one.***

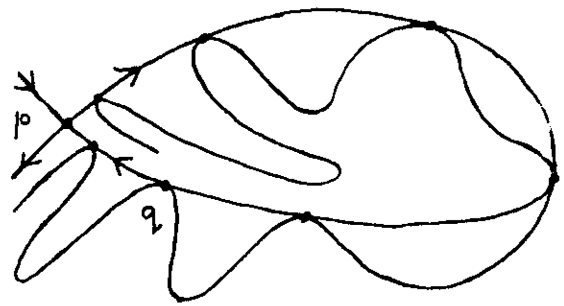
As a consequence of such unfolding the dynamics (orbit structure) of the diffeomorphisms exhibit a great number of changes (bifurcations) as the parameter evolves near the value say $\mu = 0$ corresponding to the homoclinic tangency. In particular the homoclinic tangency is accumulated by other homoclinic tangencies for values of μ approaching zero. Also many periodic points appear (or disappear) or lose hyperbolicity and change index (i.e., dimension of stable manifold). We have already seen in the last chapter that transverse homoclinic orbits imply the existence of chaotic orbits. Near homoclinic tangency we expect chaotic dynamics and even chaotic attractors, although generically there can be no nontrivial hyperbolic attractor; see Chapters VI and VII for more complete discussions, also in relation with infinitely many periodic attractors and "apparent chaos". On the other hand, there are homoclinic bifurcations where we have for most parameter values hyperbolicity: chaotic orbits but no chaotic dynamics. We believe that homoclinic tangencies are common among diffeomorphisms whose limit set is not hyperbolic; see Chapter VII.

Here, in the first three sections of this chapter, we begin to describe the bifurcation phenomena mentioned above starting with a quadratic homoclinic tangency and unfolding it so as to create transverse homoclinic orbits. And in the last section we relate this unfolding with the well known family of quadratic maps of the interval $f_\mu(y) = y^2 + \mu$.

§1. Cascades of homoclinic tangencies

Let $\phi: M \times R \to M$ be a C^3 map such that $\varphi_\mu(x) = \phi(x, \mu)$ is a diffeomorphism on M for each $\mu \in \mathbf{R}$. We shall denote such a family of diffeomorphisms simply by $\{\varphi_\mu\}$ or just φ_μ. The reason we take the family to be C^3 (and not C^1 or C^2) comes from the discussion of the period doubling bifurcation (or flip) to be presented in the next section of this chapter.

Let us start studying homoclinic tangencies and their unfoldings. Let $p = p_0$ be a hyperbolic fixed point for φ_0 and let q be a homoclinic tangency related to p, that is q is a point of tangency between $W^s(p)$ and $W^u(p)$. We assume that $W^s(p)$ and $W^u(p)$ have a quadratic (parabolic) contact at q, and just call q or its orbit $O(q)$ a quadratic homoclinic tangency.



We choose local coordinates (x_1, x_2) near q so that we can express the local components of $W^s(p)$ and $W^u(p)$ containing q by

$$W^s(p) = \{(x_1, x_2); x_2 = 0\}$$
$$W^u(p) = \{(x_1, x_2); x_2 = ax_1^2\} \qquad (1)$$

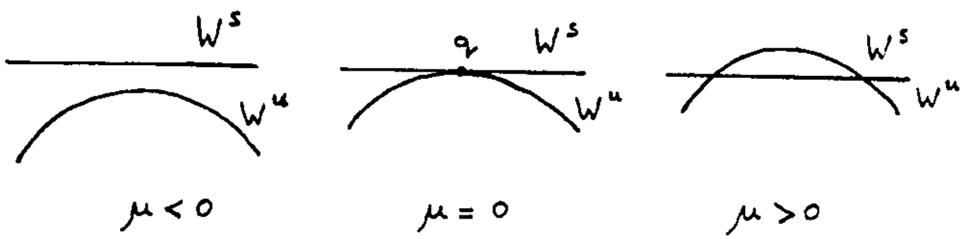
where $a \neq 0$. Since p is hyperbolic, we have for μ small a unique fixed point p_μ near p and the mapping $\mu \to p_\mu$ is differentiable (implicit function theorem). Also the local components of $W^s(p_\mu)$ and $W^u(p_\mu)$ near q depend differentiably on μ for μ near zero.

Under generic assumptions (see [**NPT**, 1983]) there are μ-dependent local coordinates such that $W^s(p_\mu)$ is given by $x_2 = 0$ and $W^u(p_\mu)$ by

$$x_2 = ax_1^2 + b\mu, \qquad a \neq 0 \quad \text{and} \quad b \neq 0 \qquad (2)$$

Then we say that the quadratic homoclinic tangency ***unfolds generically***.

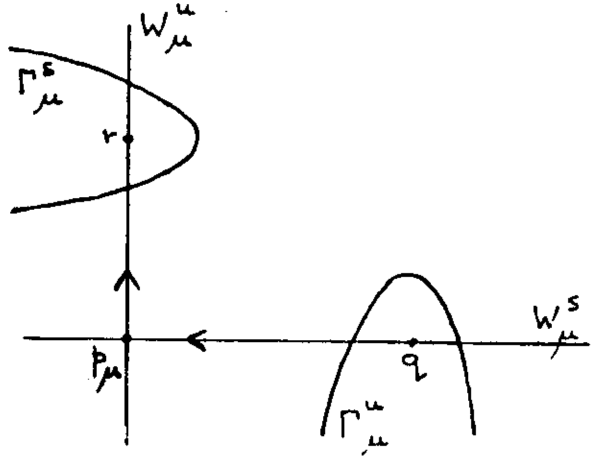
Taking $a < 0$ and $b > 0$ in (2) above, we get for the relative positions of the local components of $W^s(p_\mu)$ and $W^u(p_\mu)$:



Theorem. *Let $\{\varphi_\mu\}$ be a one-parameter family of diffeomorphisms with a quadratic homoclinic tangency q at $\mu = 0$ associated to the fixed (periodic) saddle p and suppose it unfolds generically. Then there is a sequence $\mu_n \to 0$ such that φ_{μ_n} has homoclinic tangencies $q_{\mu_n} \to q$ related to $p_{\mu_n} \to p$.*

Proof: Let $r = \varphi_0^{-N}(q)$ for some large $N > 0$ and suppose the tangency unfolds into transversal homoclinic points for $\mu > 0$. Given $\mu > 0$ near zero, there are small pieces of

"parabolas" (see the above discussion on unfoldings of homoclinic tangencies) $\Gamma_\mu^u \subset W_\mu^u$ near q and $\Gamma_\mu^s \subset W_\mu^s$ near r. We assume that, for $\mu > 0$, their position relative to W_μ^s and W_μ^u near q and r is as indicated.

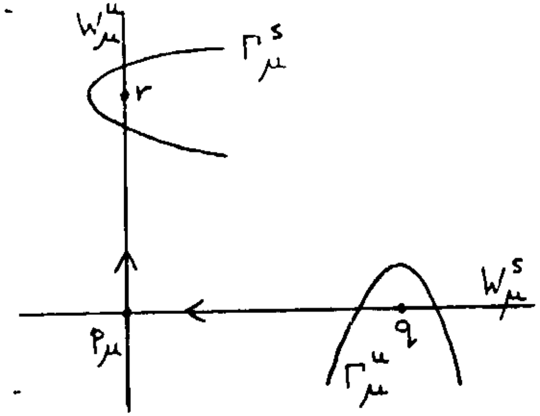


Now take $\mu = \hat{\mu}$ arbitrarily small. Clearly, if $n > 0$ is large then $\varphi_{\hat{\mu}}^{-n}(\Gamma_{\hat{\mu}}^s)$ intersects $\Gamma_{\hat{\mu}}^u$. Taking $\mu > 0$ much smaller than $\hat{\mu}$ we have, for the same integer n, that $\varphi_\mu^{-n}(\Gamma_\mu^s) \cap \Gamma_\mu^u = \phi$. Since $\varphi_\mu^{-n}(\Gamma_\mu^s)$ and Γ_μ^u depend C^3 on μ, there is some $0 < \mu_1 < \hat{\mu}$ for which $\varphi_{\mu_1}^{-n}(\Gamma_{\mu_1}^s)$ and $\Gamma_{\mu_1}^u$ are tangent say at $q_1 \in \Gamma_{\mu_1}^u$. We can repeat the argument for smaller values of $\hat{\mu}$ and so we can construct the sequences μ_n, q_n as desired, proving the result in the case indicated in the figure above. The reader can easily adapt the argument to other cases, like the one in the figure follwing the remarks below.

Remark 1: In the proof of the theorem we can take the homoclinic tangencies q_{μ_n} to be of quadratic contact: due to different curvatures, $\varphi_\mu^{-n}(\Gamma_\mu^s)$ and Γ_μ^u have a quadratic contact at their last tangency for decreasing values of μ. One can even show that these homoclinic tangencies unfold generically.

Remark 2: For the constructions in Chapter VI it is important to observe that we can even choose the values μ_n in the last theorem so that the branches of $W^s(p_{\mu_n})$ and

$W^u(p_{\mu_n})$, i.e. connected components of $W^s(p_{\mu_n}) \setminus \{p_{\mu_n}\}$ and $W^u(p_{\mu_n}) \setminus \{p_{\mu_n}\}$, which have a homoclinic tangency, also have transverse homoclinic intersections.



§2. Saddle-node and period doubling bifurcations

As we have seen in Chapter II, transverse homoclinic orbits imply the existence of horseshoes. So, near a homoclinic bifurcation we expect creation or anihilation of horseshoes. In Section 3 we describe the bifurcations due to the formation of such a horseshoe. As a preparation we recall here two of the three generic bifurcations of fixed points in one-parameter families of diffeomorphisms.

Let $\{\varphi_\mu\}_{\mu\in\mathbf{R}}$ be an arc of diffeomorphisms and x_0 a hyperbolic fixed point for φ_0. Then for μ small, φ_μ has a fixed point x_μ, called the continuation of x_0, which is near x_0 for small μ and has the same index as x_0. Thus, for x_{μ_0} to be a bifurcating orbit at least one of the two eigenvalues of $d\varphi_{\mu_0}$ at x_{μ_0} must have norm one. For generic families, we have three possible cases, as far as the eigenvalues ρ_1 and ρ_2 are concerned:

a) $\rho_1 = 1$ and $|\rho_2| < 1$ (or $|\rho_2| > 1$),

b) $\rho_1 = -1$ and $|\rho_2| < 1$ (or $|\rho_2| > 1$),

c) $\rho_1 = e^{i\theta}$, $\rho_2 = e^{-i\theta}$ for some real $\theta \neq k\pi$, $k \in \mathbf{Z}$.

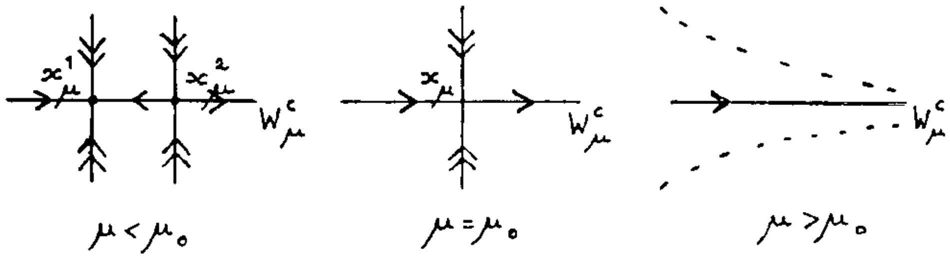
Case (c), with some further assumptions, corresponds to the so-called *Hopf bifurcation* and it will not be considered here since we will require our mappings to be dissipative (area contracting). In cases (a) and (b), there is a C^3 φ_μ-invariant curve $W^c_\mu = W^c(p_\mu)$ which is tangent at $\mu = \mu_0$ to the eigenspace associated with $\rho_1 = 1$ or $\rho_1 = -1$; it is called the centre manifold of φ_μ (see Appendix I). Thus if we let $f_\mu = \varphi_\mu | W^c_\mu$, we have the following expressions:

$$f_\mu(x) = x + ax^2 + b(\mu - \mu_0) + h\cdot o\cdot t\cdot \tag{3}$$

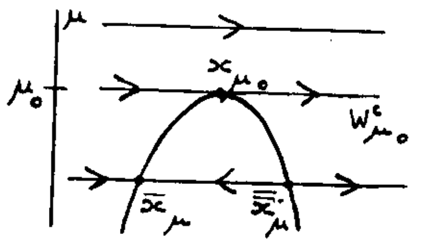
$$f_\mu(x) = -x + ax^3 + b(\mu - \mu_0)x + h\cdot o\cdot t, \tag{4}$$

$b(0) = 0$. Here (3) corresponds to the first case a) and (4) to the second case b); h.o.t stands for higher order terms. We left out the quadratic term in x in (4) because it can be removed by a change of coordinates.

In (3) we take $a \neq 0$ and call the origin a *saddle-node*. We also take $b'(0) \neq 0$ and say that the saddle-node *unfolds generically*. These conditions are clearly satisfied generically. It is easy to see from (3) that f_μ, and thus φ_μ, has two hyperbolic fixed point for $\mu < \mu_0$ and none for $\mu > \mu_0$ or vice-versa. If we consider $a > 0$, $b > 0$ and $|\rho_2| < 1$ we have the following unfolding of the saddle-node: a sink and a saddle collapse and then disappear, as shown in the figures.



The double arrows in the figures mean that the contraction in the normal direction is stronger than along W^c_μ. If we consider, for $\mu \leq \mu_0$, the curves $\mu \to \overline{x}_\mu$, $\mu \to \overline{\overline{x}}_\mu$ of fixed points, we get:

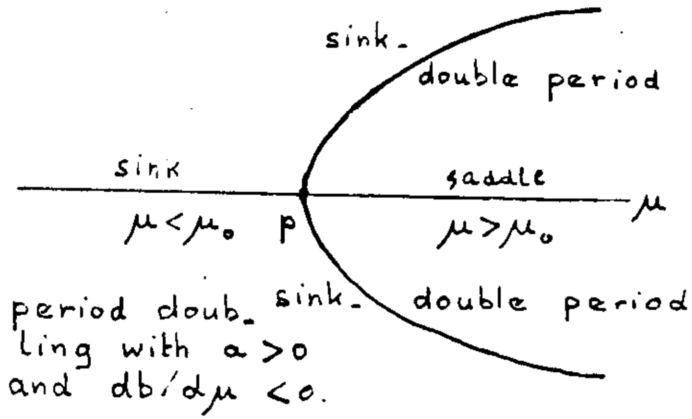


Notice that the two curves are differentiable for $|\mu - \mu_0|$ small, $\mu < \mu_0$. If we follow the curve $\mu \to \bar{\bar{x}}_\mu$ for $\mu \nearrow \mu_0$, we can then return along $\mu \to \bar{x}_\mu$ with decreasing values of μ. So the two branches can naturally be oriented as above (or vice-versa). In words: if we follow the curve of saddles for increasing values of μ, up to $\mu = \mu_0$, we then return along the curve of sinks for decreasing values of μ. This fact will play a role in the proof of the next theorem.

Now we consider the expression (4) above corresponding to the eigenvalue $\rho_1 = -1$. Similarly to what we have done before in (3), we take $a \neq 0$ (which is a generic condition) and call the orbit a *period doubling bifurcation* (or flip); we say that it *unfolds generically* if $b'(0) \neq 0$ (another generic condition!). When $a > 0$ and $b'(0) < 0$, we can easily show that there exists a unique fixed point which is a sink for $\mu < \mu_0$ and a saddle for $\mu > \mu_0$ (both with *negative eigenvalues*); for $\mu > \mu_0$ there is also a *period two* sink (with *positive eigenvalues*). Thus the name period doubling bifurcation. The results are of course similar in the other cases, where a and $b'(0)$ may have signs different from the ones above. Notice that period doubling which unfolds generically is isolated; the same is true for saddle-nodes. The assumptions and results are also similar for period doubling bifurcations of periodic orbits by just considering the power of the map equal to the period. For instance, a sink of period k may bifurcate into a saddle of period k (both with corresponding negative eigenvalues) and a sink with twice the period (and positive eigenvalues).

For the period doubling bifurcation considered above, if in the set of periodic points we identify points in the same orbit we obtain a topological 1-complex—the curve of sinks for $\mu < \mu_0$, branching off into two topological 1-manifolds: one is formed by the curve of

saddles and the other by the curve of sinks with twice the period. Notice that the sink to the left and the saddle to the right both have the same period (and a corresponding negative eigenvalue for df^k_μ, k being the period). This remark is relevant in the next section.



§3. Cascades of period doubling bifurcations and sinks

We now discuss the definitions and assumptions of the next theorem showing the existence of many sinks (or sources) and period doubling bifurcations while creating a horseshoe. The sinks that we exhibit in this chapter arise from period doubling bifurcations; they occur for different values of the parameter.

Let R be a rectangle in $\mathbf{R}^2$ and $\{\varphi_\mu\}$ a family of diffeomorphisms of R into $\mathbf{R}^2$ such that

1) $\varphi_{-1}(R) \cap (R) = \phi$,

2) $\varphi_\mu|R$ is dissipative (area contracting) for $-1 \le \mu \le 1$, that is $|\det(d\varphi_\mu)| < 1$ on R,

3) φ_1 has periodic points and they are all saddles,

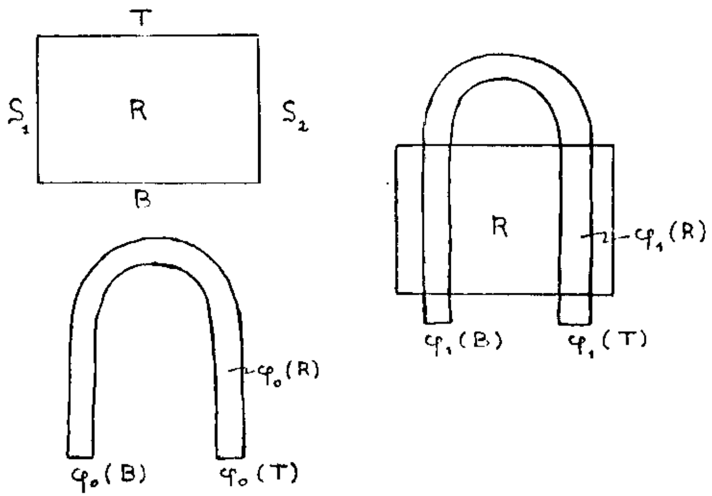
4) $\varphi_\mu(R) \cap S_1 = \phi$, $\varphi_\mu(R) \cap S_2 = \phi$, $-1 \le \mu \le 1$, where S_1, S_2 are two opposite sides in the boundary of R, say the vertical sides,

5) $\varphi_\mu(T) \cap R = \phi$, $\varphi_\mu(B) \cap R = \phi$, $-1 \le \mu \le 1$, where T is the top side of R and B is the bottom side.

In this section we also assume the following ***generic*** (***residual*** or ***Baire second category***) condition on the family $\{\varphi_\mu\}$.

6) φ_μ has at most one nonhyperbolic periodic orbit for each $-1 \le \mu \le 1$ and this orbit must correspond either to a saddle-node or to a period doubling bifurcation which unfolds generically. (Because φ_μ is area contracting there is no Hopf bifurcation).

Although we did not formally require φ_1 to be a horseshoe mapping like in Chapter II, that is precisely the situation we have in mind. In this case, we say that we have an area decreasing family creating a horseshoe, as in the following figures.

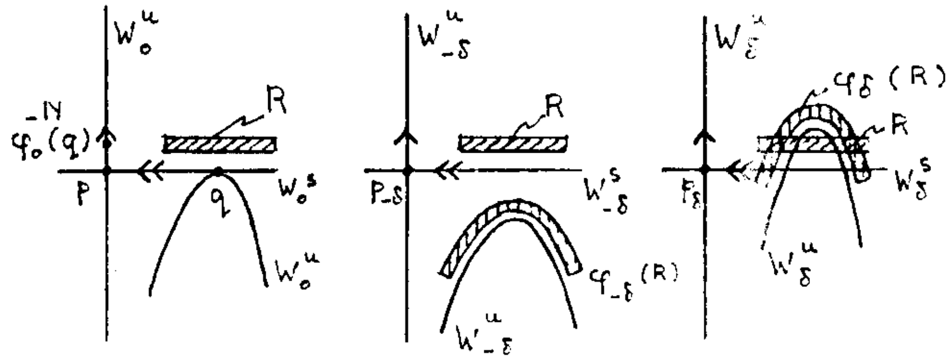


Before continuing the discussion, we want to point out that the above conditions are satisfied in a generically unfolding homoclinic tangency replacing φ_μ by φ_μ^N, taking $-\delta < \mu < +\delta$ instead of $-1 < \mu < +1$ and choosing R appropriately. Of course, to get the area decreasing property, we assume the determinant of the Jacobian of the map at the fixed (or periodic) saddle with a homoclinic tangency to be less than one in absolute value. To see the creation of a horseshoe, let φ_μ be such that

(i) φ_0 has a fixed saddle p and $|\det(d\varphi_0)_p| < 1$;

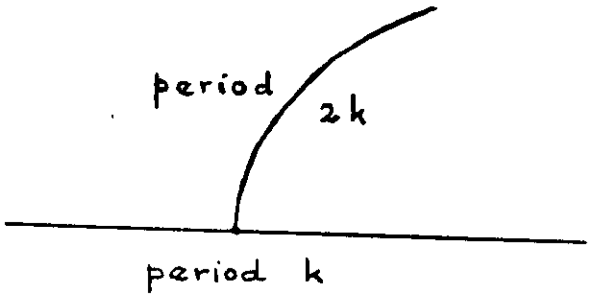
(ii) there is a generically unfolding homoclinic tangency q associate to p.

We then claim that for each neighbourhood V of q there exists a rectangle $R \subset V$, a number $\delta > 0$ and an integer $N > 0$ such that $\varphi_\mu^N|R$ creates a horseshoe for $-\delta < \mu < \delta$: take R to be a thin rectangle near q and parallel to the local component of W_μ^s as in the figure.



Let us see why we can choose R, δ and N as wished. First, for μ small we choose C^1 coordinates linearizing each φ_μ in a fixed neighbourhood of p containing an arc $\ell^s \subset W_0^s$ from p to q; these coordinates may be chosen to depend continuously on μ (see Appendix I). We then choose R to be thin and sufficiently close to W_0^s so that its projection on W_0^u parallel to W_0^s contains in its interior $\varphi_0^N(q)$ for some large N. Then $\varphi_\mu^N(R)$ will be a "curved box" close to an arc in W_0^u near q. For μ near zero, we then have the situation indicated in the figure. One can then apply arguments similar to those in Chapter II to show that $\varphi_\mu^N|R$ is area decreasing for $-\delta \le \mu \le \delta$ and that $\varphi_\delta^N|R$ has its maximal invariant set hyperbolic with dense subset of periodic orbits; see also Section 4. In fact, we observe that although the configuration R, $\varphi_\delta^N(R)$ resembles the situation in Chapter II, the rectangles considered are quite different: there we had a long rectangle containing p and q; here the rectangle is contained in a small neighbourhood of q.

We now return to the general discussion about creating horseshoes. For $\varphi_\mu: \mathbf{R}^2 \to \mathbf{R}^2$ as before, let Per (φ_μ) be the set of periodic orbits of φ_μ and $P = \{(x,\mu); x \in \text{ Per } (\varphi_\mu)\}$. We now define the topological space $\tilde{P} = P/\sim$, where the equivalence relation $\sim$ is the identification of points in the same orbit. A component of $\tilde{P}$ through $(O(x),\mu), O(x)$ being the orbit of the periodic point x, is a continuous curve except at period doubling or undoubling bifurcations where it branches and looks like

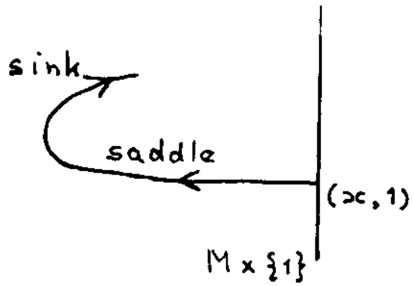


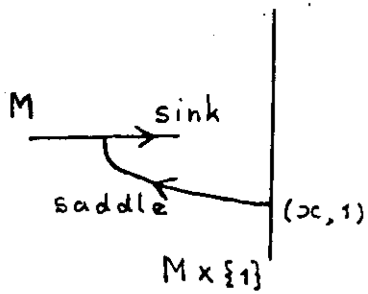
Theorem [YA, 1983]. *Let $\varphi_\mu: \mathbf{R}^2 \to \mathbf{R}^2$ be a family of orientation preserving diffeomorphisms satisfying conditions (1) to (6). Then each $(0(x),1) \in \tilde{P}$ has a component containing attracting periodic orbits (sinks) of periodic $2^n k$ for each $n \ge 0$, where k is the period of x for φ_1.*

Remark: For orientation reversing families $\{\varphi_\mu\}$ one just considers the squares $\{\varphi_\mu^2\}$; if they satisfy the conditions (1) to (6) we get a corresponding result.

Proof: Let $(0(x),1) \in \tilde{P}$ and assume first that $(d\varphi_1^k)_x$ has ***positive eigenvalues***, k being the period of x. By the implicit function theorem, there is a (unique) continuous path Γ in $\tilde{P}$ through $(0(x),1)$ which we follow for decreasing values of μ. We then must reach a

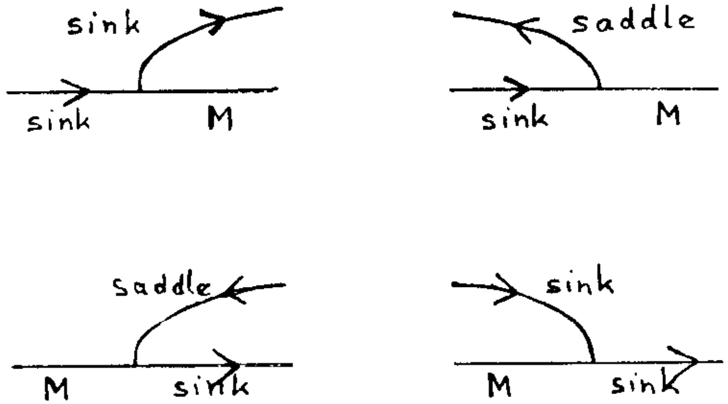
bifurcation for otherwise we could follow Γ up to $M \times \{-1\}$ (strictly speaking, Γ is not a curve in $M \times [-1,+1]$ but in $\tilde{P}$, the space of periodic orbits in $M \times [-1,+1]$). In fact, by conditions (4) and (5), the maximal invariant set of φ_μ in R is bounded away from ∂R (periodic points cannot escape through ∂R) and also we cannot terminate Γ in $R \times (-1,1)$ because we can always prolong a path of saddle points. But φ_{-1} has no periodic points in R and so we must reach a bifurcation point and this must correspond to a saddle-node or a period undoubling bifurcation. In both cases we then follow the path of sinks that emanates from the bifurcating orbit for increasing values of μ (see discurssions before on saddle-nodes and period doubling bifurcations):



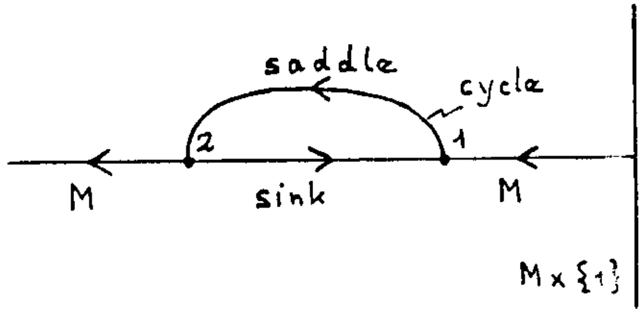


In what follows we always prolong Γ at a period doubling or undoubling bifurcation avoiding paths of saddles with corresponding negative eigenvalues (***Moebius paths***) and orient the path positvely (for increasing values of μ) if it is a path of sinks and negatively if it is a path of saddles. See the corresponding figures below clarifying the convention. At saddle-nodes the paths are oriented following the same convention. Now we cannot reach back $M \times \{1\}$ since φ_1 has no sinks. Also we cannot have a *cycle*, i.e. Γ cannot return to itself since in each bifurcating point there is one path of saddles and one path of sinks (Moebius paths are not counted here).

We also claim we cannot terminate Γ in $M \times (-1,1)$ if we go through only finitely many bifurcating orbits or even infinitely many ones say (x_i,μ_i) with bounded periods. In fact, in the first case, from our discussions on hyperbolic and generic bifurcating periodic orbits, we could clearly prolongate Γ. In the second case we can consider a limiting point $(\tilde{x},\mu)$ of (x_i,μ_i) and argue that $(\tilde{x},\mu) \in P$ and then by the genericity assumption (6) on φ_μ, $(\bar{x}, \mu)$ had to be locally isolated as a bifurcating orbit of bounded period which is not the case. Thus we must go through infinitely many bifurcating periodic orbits with unbounded periods. This can be achieved only if we go through infinitely many period doubling bifurcations with unbounded periods, which then clearly implies the result in this case where we started the path at a saddle $(0(x), 1) \in \tilde{P}$ with positive eigenvalues.



Let us now begin with $(0(x), 1) \in \tilde{P}$ such that the eigenvalues of $d\varphi_1^k(x)$ are negative, where k is the period of x, and a path through it in $\tilde{P}$, i.e. a Moebius path. We will show that the result is also true in this case. Before we do that, let us again orient in the positive μ-direction paths of sinks and in the negative μ-direction both paths of saddles with positive eigenvalues (which we just call path of saddles) and Moebius paths. So let Γ be a Moebius path starting at $(0(x), 1)$. As argued before, Γ must go through bifurcating orbits and the first one must be a period doubling bifurcation. We then follow the path of saddles of twice the period that emanates from it. At the next bifurcating orbit we repeat the procedure of prolonging Γ along the unique non Moebius path emanating from it. But already at this point we may get a cycle! That is, a closed oriented path of periodic orbits not containing any Moebius curves. The figure below illustrates this possibility.



If no such cycle appears in Γ we are done. In order to solve the problem with cycles we observe that in each cycle the number of period doublings equals the number of period undoublings. At each period doubling there is an incomming Moebius branch and at each period undoubling there is an outgoing Moebius branch. We then identify in $\tilde{P}$, the space of periodic orbits in $M \times [-1, +1]$, each cycle to one point. The points have now the same (finite) number of ingoing Moebius paths as outgoing Moebius paths. Next we make in this reduced space ($\tilde{P}$ modulo cycles) a path $\tilde{\Gamma}$, starting at $(0(x), 1)$ and consisting entirely of Moebius paths such that:

- the orientation of $\tilde{\Gamma}$ agrees with the orientations of the Moebius paths it follows;
- $\tilde{\Gamma}$ passes along each Moebius path at most once.

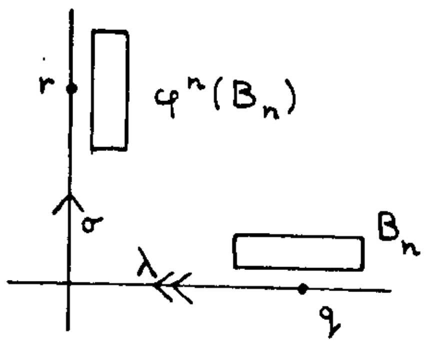
Such a path can always be continued whenever it reaches a collapsed cycle (the number of ingoing branches equals the number of outgoing branches); such a path cannot end in a periodic orbit in $M \times \{1\}$ since all Moebius paths are oriented towards lower values of μ. So, there are two possibilities:

either $\tilde{\Gamma}$ is finite, but then it terminates in a period doubling point not belonging to a cycle. As we observed before, then we are done,

or, $\tilde{\Gamma}$ is infinite. In this case the path $\tilde{\Gamma}$ induces a connected graph Γ in $\tilde{P}$, consisting of "arcs of $\tilde{\Gamma}$" and those cycles where $\tilde{\Gamma}$ passes through. Clearly Γ is infinite and hence contains an infinite number of bifurcating periodic orbits. Also now we are back to a previous case. ∎

§4. Homoclinic tangencies, scaling and quadratic maps

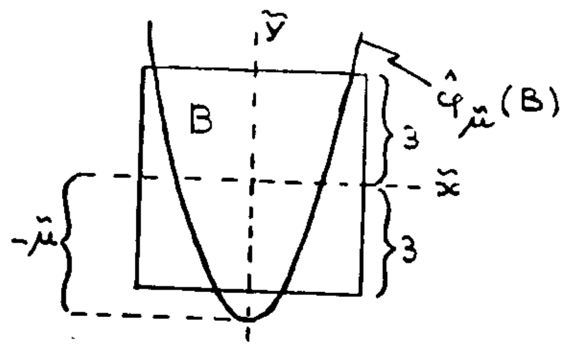
We consider a one parameter family of diffeomorphisms $\varphi_\mu: M \to M$, M a 2-manifold, which has for $\mu = 0$ a homoclinic tangency. Let p_μ denote the saddle point of φ_μ which is related, for $\mu = 0$, to this tangency. We assume the tangency of $W^u(p_0)$ and $W^s(p_0)$ to, be generic (parabolic contact) and also to unfold generically.



Before we go into technicalities, we want to give a heuristic idea of the construction to be described in this section, and its consequences. Near p_0 we take linearizing coordinates x, y so that $\varphi_0(x,y) = (\lambda \cdot x, \sigma \cdot y)$ with $0 < |\lambda| < 1 < |\sigma|$. We assume λ and σ to be positive (otherwise we replace φ_μ by φ_μ^2), and that $\lambda \cdot \sigma < 1$ (if $\lambda \cdot \sigma > 1$ we replace φ_μ by φ_μ^{-1} and if $\lambda \cdot \sigma = 1$ our construction does not work). Let q and r be points on the orbit of tangency in the domain of the linearizing coordinates as indicated. So, for some N, $\varphi_0^N(r) = q$. For each sufficiently big n, we take a box B_n near q such that $\varphi_0^n(B_n)$ is a box near r as indicated. We consider $\varphi_\mu^{n+N}(B_n)$, and especially its position relative to B_n. As was already mentioned in Section 3 of this chapter, if one chooses B_n carefully, then, for n sufficiently big, $\varphi_\mu^{n+N}(B_n)$ will cross over B_n so as to create a horseshoe. We shall not only prove this but even show that, after applying n-dependent coordinate transformations to

both the x, y variables and the parmeter μ (denoting the new variables by $\tilde{x}, \tilde{y}$, and $\tilde{\mu}$), $\varphi_{\tilde{\mu}}^{n+N}$ converges for $n \to \infty$ to the map $\hat{\varphi}_{\tilde{\mu}}$, given by $\hat{\varphi}_{\tilde{\mu}}(\tilde{x}, \tilde{y}) = (\tilde{y}, \tilde{y}^2 + \tilde{\mu})$.

Taking the box B_n in these (n-dependent) coordinates equal to $B = \{(\tilde{x}, \tilde{y}) | |\tilde{x}| \le 3, |\tilde{y}| \le 3\}$ we get the horseshoe formation when $\tilde{\mu}$ decreases from say 4 to -4, at least for n sufficiently big.



Note that this limiting map is not a diffeomorphism any more. This is related to the fact that φ_μ is area contracting at p_μ and hence φ_μ^{n+N}, for $n \to \infty$, becomes more and more area contracting, $\varphi_\mu^{n+N}(B_n)$ tending to be just a curve.

For the limiting map, the value $\tilde{x}$ is unimportant. Restricting to the $\tilde{y}$ variable we have

$$\tilde{y} \mapsto \tilde{y}^2 + \tilde{\mu},$$

which is the well known one-parameter family of quadratic one-dimensional maps, that has been studied e.g. in [CE, 1980].

This being the limiting map, φ_μ^{n+N} "contains" approximations of this family of quadratic maps and hence exhibits much of its complexity, see [S,1981]: hyperbolic sets, period doubling bifurcations and all other phenomena that are persistent under C^r ($r \ge 2$) perturbations.

Because it will be used later, we give here one example of extending a fact about quadratic maps to the one-parameter families like φ_μ. For $\tilde{\mu}$ near zero, the map $\tilde{y} \mapsto \tilde{y}^2 + \tilde{\mu}$ has an attracting fixed point near zero. Let $\mu_n \to 0$ be the sequence of μ-values corresponding to $\tilde{\mu} = 0$ in the different reparametrizations of the μ-variable. Then for n sufficiently big and μ near μ_n, φ_μ^{n+N} has an attracting fixed point.

Now we give a more formal and complete description of the result. First we have to state some extra assumptions on the 1-parameter family φ_μ. As we mentioned already we assume that the eigenvalues λ, σ of $(d\varphi_0)_{p_0}$ are positive and satisfy $\lambda \cdot \sigma < 1$. Also we need C^2 linearizing coordinates of φ_μ near p_μ. For this reason we require $\varphi_\mu(x,y)$ to be C^∞ in (μ, x, y). The C^2 linearizing coordinates (*μ-dependent*) then exist, provided that some generic (even open and dense) conditions are satisfied by the eigenvalues λ and σ (see [S,1958]); they depend continuously, in the C^2 topology, on μ.

Theorem. *For a one-parameter family φ_μ as above, with q a point on the orbit of tangency for $\mu = 0$, there are a constant N and, for each positive integer n, reparametrizations $\mu = M_n(\tilde{\mu})$ of the μ variable and $\tilde{\mu}$-dependent coordinate transformations*

$$(\tilde{x}, \tilde{y}) \mapsto (x, y) = \Psi_{n,\tilde{\mu}}(\tilde{x}, \tilde{y})$$

such that:

- *for each compact set K in the $\tilde{\mu}, \tilde{x}, \tilde{y}$ space the images of K under the maps*

$$(\tilde{\mu}, \tilde{x}, \tilde{y}) \mapsto (M_n(\tilde{\mu}), \Psi_{n,\tilde{\mu}}(\tilde{x}, \tilde{y}))$$

converge, for $n \to \infty$, in the μ, x, y space, to $(0, q)$;

- *the domains of the maps*

$$(\tilde{\mu}, \tilde{x}, \tilde{y}) \mapsto (\tilde{\mu}, (\Psi_{n,\tilde{\mu}}^{-1} \circ \varphi_{M_n(\tilde{\mu})}^{n+N} \circ \Psi_{n,\tilde{\mu}}))$$

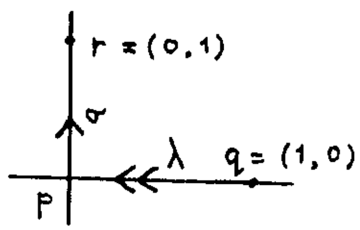
converge, for $n \to \infty$, to all of $\mathbf{R}^3$, and the maps converge in the C^2 topology, to the map

$$(\tilde{\mu}, \tilde{x}, \tilde{y}) \mapsto (\tilde{\mu}, \tilde{\varphi}_{\tilde{\mu}}(\tilde{x}, \tilde{y}))$$

with $\tilde{\varphi}_{\tilde{\mu}}(\tilde{x}, \tilde{y}) = (\tilde{y}, \tilde{y}^2 + \tilde{\mu})$.

This theorem is an expanded version of a remark in §6.7 (p.336) of [GH,1983]. *Actually, if there are C^k linearizing coordinates, $k \geq 2$, the convergence to $\tilde{\varphi}_{\tilde{\mu}}$ is in the C^k topology.* This is a consequence of our proof.

Proof: We start by choosing μ-dependent C^2 linearizing coordinates near p_μ. We denote them by (x, y). For $\mu = 0$ we have $\varphi_0(x,y) = (\lambda x, \sigma y)$ with $0 < \lambda < 1 < \sigma$ and $\lambda \cdot \sigma < 1$. Let q be a point in the orbit of tangency in the "local" stable manifold of p and r such a point in the "local" unstable manifold of p.



By multiplying x, y with constants, we arrange that $q = (1,0)$ and $r = (0,1)$. Since both r and q are on the orbit of tangency, there is N such that $\varphi_0^N(r) = q$. For μ near zero we adapt our linearizing coordinates so that:

- $\varphi_\mu^N(0,1)$ is a local maximum of the y-coordinate restricted to $W^u(p_\mu)$;
- the x-coordinate of $\varphi_\mu^N(0,1)$ is 1.

We reparametrize μ is such a way that the y-coordinate of $\varphi_\mu^N(0,1)$ is μ.

After these preliminary steps we can write φ_μ^N, near (0,1), as

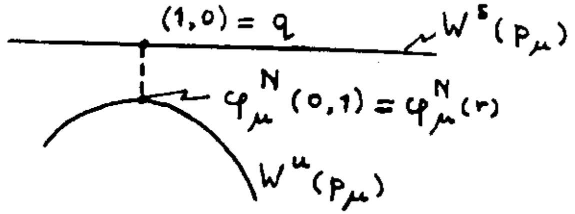
$$(x, 1+y) \mapsto (1,0) + (H_1(\mu,x,y), H_2(\mu,x,y))$$

with

$$H_1(\mu,x,y) = \alpha \cdot y + \tilde{H}_1(\mu,x,y),$$
$$H_2(\mu,x,y) = \beta \cdot y^2 + \mu + \gamma \cdot x + \tilde{H}_2(\mu,x,y)),$$

where α, β, γ are non-zero constants, and where, for $\mu = x = y = 0$:

$$(1) \qquad \begin{cases} \tilde{H}_1 = \partial_y \tilde{H}_1 = \partial_\mu \tilde{H}_1 = 0, \\ \tilde{H}_2 = \partial_x \tilde{H}_2 = \partial_y \tilde{H}_2 = \partial_\mu \tilde{H}_2 = \partial_{yy} \tilde{H}_2 = \partial_{y\mu} \tilde{H}_2 = \partial_{\mu\mu} \tilde{H}_2 = 0. \end{cases}$$



The functions H_i and $\tilde{H}_i$ are clearly C^2 since φ_μ is C^∞ and the x,y-coordinates are C^2 in x,y; in fact, instead of $\tilde{H}_2 = \partial_\mu \tilde{H}_2 = \partial_{\mu\mu}\tilde{H}_2 = 0$ and $\partial_y \tilde{H}_2 = \partial_{\mu y}\tilde{H}_2 = 0$, we should put $\tilde{H}_2(\mu,0,0) \equiv 0$ and $\partial_y \tilde{H}_2(\mu,0,0) \equiv 0$.

Next we define an *n-dependent* reparametrization of μ and a μ-dependent coordinate transformation by the following formulas:

$$\mu = \sigma^{-2n}\cdot\overline{\mu} - \gamma\cdot\lambda^n + \sigma^{-n} \qquad \overline{\mu} = \sigma^{2n}\cdot\mu + \gamma\cdot\lambda^n\cdot\sigma^{2n} - \sigma^n$$

$$x = 1 + \sigma^{-n}\cdot\overline{x} \qquad \overline{x} = \sigma^n\cdot(x-1)$$

$$y = \sigma^{-n} + \sigma^{-2n}\cdot\overline{y} \qquad \overline{y} = \sigma^{2n}\cdot y - \sigma^n.$$

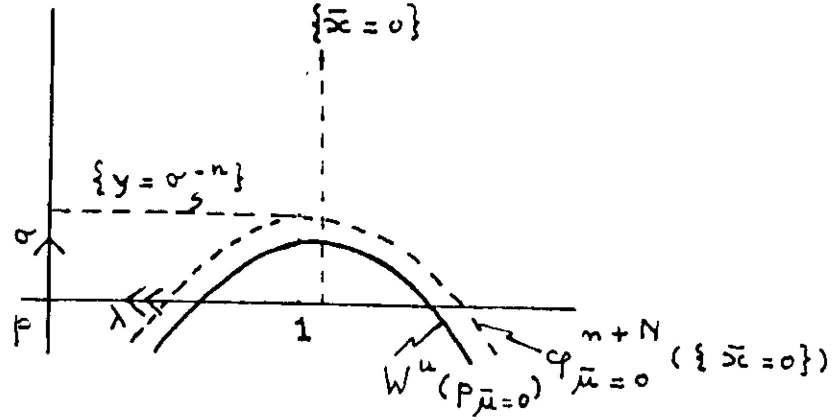
Note that these are *not* yet the final reparametrizations and coordinate transformations but the final ones differ from these just by constant factors, as made explicit at the end of the proof. Also. σ and λ depend on μ and hence on $\overline{\mu}$ although this is not expressed in the above formulas. Finally, note that a fixed box in the $\overline{x},\overline{y}$ coordinates, gives for $n \to \infty$ boxes converging to q in the x,y coordinates.

We now start with our main calculation: expressing φ_μ^{n+N} in terms of $\overline{\mu}, \overline{x}$ and $\overline{y}$. Let $(\overline{\mu},\overline{x},\overline{y})$ denote a point. The (μ,x,y) variables of this point are (see above):

$$\mu = \sigma^{-2n}\cdot\overline{\mu} - \gamma\cdot\lambda^n + \sigma^{-n}, \quad x = 1 + \sigma^{-n}\cdot\overline{x}, \quad y = \sigma^{-n} + \sigma^{-2n}\cdot\overline{y}.$$

After applying φ_μ^n to this point we get as x,y coordinates (μ does not change):

$$x = \lambda^n\cdot(1+\sigma^{-n}\overline{x}), \qquad y = 1 + \sigma^{-n}\cdot\overline{y}.$$



Next we apply φ_μ^N and find;

$$x = 1 + \alpha\cdot\sigma^{-n}\cdot\overline{y} + \tilde{H}_1(\mu, \lambda^n\cdot(1+\sigma^{-n}\overline{x}), \sigma^{-n}\cdot\overline{y})$$

$$y = \beta\cdot\sigma^{-2n}\cdot\overline{y}^2 + (\sigma^{-2n}\overline{\mu} - \gamma\cdot\lambda^n + \sigma^{-n})$$
$$+\gamma\cdot\lambda^n\cdot(1+\sigma^{-n}\overline{x}) + \tilde{H}_2(\mu, \lambda^n\cdot(1+\sigma^{-n}), \sigma^{-n}\cdot\overline{y}).$$

Transforming this back to the $\overline{x},\overline{y}$ coordinates, and denoting the values of these coordinates of the new point by $\overline{\overline{x}},\overline{\overline{y}}$ we have:

(2)

$$\begin{cases} \overline{\overline{x}} = \alpha\overline{y} + \sigma^n\cdot\tilde{H}_1(\sigma^{-2n}\cdot\overline{\mu} - \gamma\cdot\lambda^n + \sigma^{-n}, \lambda^n\cdot(1+\sigma^{-n}\overline{x}), \sigma^{-n}\cdot\overline{y}) \\ \overline{\overline{y}} = \beta\overline{y}^2 + \overline{\mu} + \gamma\cdot\lambda^n\cdot\sigma^n\cdot\overline{x} + \sigma^{2n}\cdot\tilde{H}_2(\sigma^{-2n}\cdot\overline{\mu} - \gamma\cdot\lambda^n + \sigma^{-n}, \lambda\cdot(1+\sigma^{-n}\overline{x}), \sigma^{-n}\cdot\overline{y}). \end{cases}$$

Next, we need to show that in the above expression certain parts converge to zero for $n \to \infty$ in the C^2 topology (uniformly on compacta in the $\overline{\mu},\overline{x},\overline{y}$ coordinates).

In the expression for $\overline{\overline{y}}$, the term $\gamma\cdot\lambda^n\cdot\sigma^n\cdot\overline{x}$ goes clearly to zero because $\lambda\sigma < 1$. The terms involving $\tilde{H}_i$ are more complicated. We first observe that when

$$(\overline{\mu},\overline{x},\overline{y})$$

remains bounded, the corresponding values of

$$(\mu, x, (y-1))$$

which are substituted in $\tilde{H}_1$ satisfy:

$$(3)\qquad \begin{cases} \mu = O(\sigma^{-n}) \\ x = O(\lambda^n) \\ y - 1 = O(\sigma^{-n}) \end{cases}$$

as n goes to infinite. Next we define

$$\overline{H}_1(\overline{\mu},\overline{x},\overline{y}) = \sigma^n \cdot \tilde{H}_1(\sigma^{-2n}\cdot\overline{\mu} - \gamma\cdot\lambda^n + \sigma^{-n}, \lambda^n\cdot(1+\sigma^{-n}\cdot\overline{x}), \sigma^{-n}\cdot\overline{y}).$$

Then

$$\overline{H}_1(0,0,0) = \sigma^n\cdot\tilde{H}_1(-\gamma\cdot\lambda^n+\sigma^{-n},\lambda^n,0) = \sigma^n\cdot(0(\lambda^n)+0(\lambda^{2n})),$$

which converges to zero for $n \to \infty$. Next, the first and second order derivatives of $\overline{H}_1(\overline{\mu},\overline{x},\overline{y})$ converge to zero (uniformly on compacta); this follows from (1), (3) and $0 < \lambda\sigma < 1$. In fact, the derivatives of $\overline{H}_1$ are easier to estimate than $\overline{H}_1$ itself. This is the way in which one proves that $\overline{H}_1$ goes to zero as announced. The same procedure works for the corresponding expression in the formula for $\overline{\overline{y}}$.

So, for $n \to \infty$, the transformation formulas (2) converge to

$$\begin{pmatrix}\overline{x}\\ \overline{y}\end{pmatrix} \mapsto \begin{pmatrix}\alpha\overline{y}\\ \beta\overline{y}^{2}+\overline{\mu}\end{pmatrix}.$$

By the substitution

$$\overline{\mu} = \beta^{-1}\tilde{\mu}$$

$$\overline{x} = \alpha\beta^{-1}\tilde{x}$$

$$\overline{y} = \beta^{-1}\tilde{y},$$

this limiting transformation becomes

$$\begin{pmatrix}\tilde{x}\\ \tilde{y}\end{pmatrix} \mapsto \begin{pmatrix}\tilde{y}\\ \tilde{y}^{2}+\tilde{\mu}\end{pmatrix}.$$

Now the theorem is proved: we have the announced transformation as limit of φ_μ^{n+N}, composed with suitable coordinate transformations and reparametrizations of μ. ∎

We observe that the calculation above, under stronger hypothesis, was carried out independently in [**TY**,1986].

Remark 1: One can also consider the effect of our (rescaling) coordinate transformations on the stable and unstable foliations. If $\mathcal{F}_\mu^s$ and $\mathcal{F}_\mu^u$ are stable and unstable foliations, defined in a neighbourhood of p_μ (i.e. $W^s(p_\mu)$ is a leaf of $\mathcal{F}_\mu^s$, φ_μ maps leaves of $\mathcal{F}_\mu^s$ into such leaves or maps them outside the domain of definition of $\mathcal{F}_\mu^s$, and the tangent directions of $\mathcal{F}_\mu^s$ depend continuously on x, y, and μ; similarly for $\mathcal{F}_\mu^u$), then we can extend $\mathcal{F}_\mu^s$ by negative iterates of φ_μ and $\mathcal{F}_\mu^u$ by positive iterates of φ_μ till the domains of definition of both these foliations contain the point of tangency of $W^u(p_0)$ and $W^s(p_0)$ in their interior. Next we consider these foliations with respect to the $\overline{x},\overline{y},\overline{\mu}$ - or the $\tilde{x},\tilde{y},\tilde{\mu}$ - coordinates (depending on n). It turns out, by estimates which are much like the above estimates, that in these coordinates $\mathcal{F}_\mu^s$ and $\mathcal{F}_\mu^u$ converge for $n \to \infty$. In the $\tilde{x},\tilde{y},\tilde{\mu}$ coordinates, the limiting leaves of $\mathcal{F}_{\tilde{\mu}}^u$ consist of horizontal curves, the limiting leaves of $\mathcal{F}_{\tilde{\mu}}^u$ consist of the parabolas $\{\tilde{y} = \tilde{x}^2 + a | a \in \mathbf{R}\}$.

Remark 2: If the maps $\varphi_{M_n(\tilde{\mu})}^{n+N}$ are not orientation preserving, then we may replace them by the square $\varphi_{M_n(\tilde{\mu})}^{2(n+N)}$ which is orientable: after reparametrizing according to $\Psi_{n,\tilde{\mu}}$, these maps converge to $(\hat{\varphi}_{\tilde{\mu}})^2$, and, due to renormalization, we know that $(\hat{\varphi}_{\tilde{\mu}})^2$ has, after rescaling $\tilde{\mu}$ and $\overline{x}$, the same properties as $\hat{\varphi}_{\tilde{\mu}}$.

CHAPTER IV

CANTOR SETS IN DYNAMICS AND FRACTAL DIMENSIONS

As already indicated in earlier chapters, the closure of a set of homoclinic intersections is often a Cantor set. In the following chapters, concerning the study of homoclinic bifurcations, we shall have to impose, in the formulations of several results, conditions on such Cantor sets. These conditions will involve numerical invariants like Hausdorff dimension, limit capacity, (local) thickness and denseness, which we discuss in this chapter. When these invariants are non integers, one speaks of sets of *fractal dimensions*; often nowadays the name *fractal* is associated to sets whose topological dimension is smaller than their Hausdorff dimension, like the Cantor sets we deal with here.

Since these Cantor sets are not of the most general type, we begin our discussion in the present chapter with the description of "dynamically defined (or regular) Cantor sets"; they form the class of Cantor sets in which we are mainly interested in Dynamical Systems. We prove several results concerning the relations between Hausdorff dimension, limit capacity and (local) thickness and denseness and show that they vary continuously with the maps defining the Cantor sets. These Cantor sets have Hausdorff dimension smaller than one, and thus their Lebesgue measure is zero. This is not in general the case when the surface diffeomorphism giving rise to the Cantor set is of class C^1, as shown by a counterexample due to Bowen and presented here. Apart from these results that are needed in the next chapters, we derive some additional information concerning the regularity of these Cantor sets, which in our view is of independent interest and bound to play a role in Dynamics.

We observe that much of the results in this chapter concern Cantor sets that arise from hyperbolic invariant sets of 2-dimensional diffeomorphisms. For hyperbolic sets in higher dimensions, our present knowledge is rather more limited: it is not even known that the definition of local Hausdorff dimension is independente of the initial point.

§1. Dynamically defined Cantor sets

Let $\varphi : M \to M$ be a C^3-diffeomorphism of a 2-manifold M with a hyperbolic fixed point p of saddle type which is part of a (non-trivial) basic set Λ. (*Actually, we will see that it is enough to take φ of class C^2; see the remark at the end of this section*). By a basic set we mean a compact hyperbolic invariant set with a dense orbit, whose periodic orbits are dense and which is the maximal invariant set in a neighbourhood of it. Non-trivial here means that it does not consist of (finitely many) periodic orbits. As an example, one may think of the "maximal invariant subset of R" related to a transversal homoclinic orbit, as analysed in Chapter II. For $p \in \Lambda$, the subset $\Lambda \cap W^s(p)$ of $W^s(p)$ is what we want to call a dynamically defined Cantor set. To be more precise, let $\alpha : \mathbf{R} \to W^s(p)$ be a smooth identification, such that $\alpha^{-1} \circ (\varphi | W^s(p)) \circ \alpha$ is a linear contraction (see [S. 1957]). Let K be an open and compact neighbourhoord of 0 in $\alpha^{-1}(W^s(p) \cap \Lambda)$; K is called a ***dynamically defined Cantor set***. Usually, we even assume that K is obtained by intersecting $\alpha^{-1}(W^s(p) \cap \Lambda)$ with an interval $K_0 \subset R$, containing 0, and whose boundary points are not contained in $\alpha^{-1}(W^s(p) \cap \Lambda)$. We discuss some of the main properties of these Cantor sets.

Scaling: If $\alpha^{-1} \circ (\varphi | W^s(p)) \circ \alpha$ is a linear contraction by λ, which we assume positive, then, since Λ is invariant under φ,

$$\lambda \cdot K = K \cap (\lambda \cdot K_0)$$

where, for $A \subset \mathbf{R}$ and $\lambda \in \mathbf{R}$, $\lambda \cdot A = \{\lambda \cdot a \mid a \in A\}$. This means that the choice of the interval K_0 is not very essential: in each interval $[\lambda \cdot a, a]$ one has the same geometry.

Expanding structure: There is a smooth expanding map $\Psi : K \to K$ with some remarkable properties. We first construct this map. As in Chapter II we choose an unstable foliation $\mathcal{F}^u$, defined on a neighbourhood U of Λ. Since the diffeomorphism φ is C^3, this foliation is $C^{1+\epsilon}$. If the interval K_0 is sufficiently big, then we have a projection π, along leaves of $\mathcal{F}^u$, of a neighbourhood U' of Λ to $\alpha(K_0)$; clearly $\pi(\Lambda) = \alpha(K)$. This projection is in general not unique: one leaf of $\mathcal{F}^u$ may have more than one intersection with $\alpha(K_0)$. Since Λ is totally disconnected, one can still take π, on a small neighbourhood of Λ, continuous and hence differentiable (in fact $C^{1+\epsilon}$). The derivative of $\pi|(W^s(p) \cap U')$ is bounded and bounded away from zero since the components of $W^s(p) \cap U'$ are leaves of the stable foliation, which is transverse to $\mathcal{F}^u$. For N sufficiently big,

$$\Psi = \alpha^{-1} \circ \pi \circ \varphi^{-N} \circ \alpha : K_0 \to K_0$$

is, where defined, expanding in the sense that the derivative has *norm bigger than one*. Indeed, the possible contractions in $\pi|(W^s(p) \cap U')$ are compensated by φ^{-N}. From the above construction it follows that $\Psi(K) = K$ (we also denote $\Psi|K$ by Ψ) and that Ψ is $C^{1+\epsilon}$ on a neighbourhood of K.

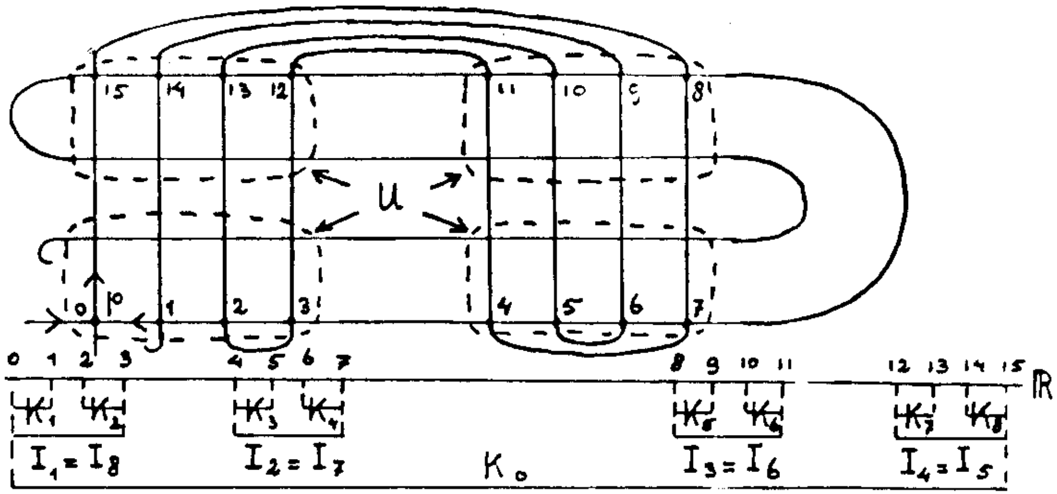
Our assumption that K_0 has to be sufficiently big is no real restriction due to the scaling property. The non-uniqueness of Ψ, due to the non-uniqueness of π is still a problem, but this can be bypassed as a consequence of our construction of Markov partitions in Appendix II.

Markov partitions: For a Cantor set K and an expanding map Ψ as above, we define a Markov partition as a finite set of disjoint intervals $K_1, \ldots, K_k \subset K_0$ such that

- Ψ is defined on a neighbourhood of each K_i, $i \geq 1$;
- K is contained in $\bigcup_{i=1}^{k} K_i$, and the boundary of each K_i is contained in K;
- for each $1 \leq i \leq k$, $\Psi(K_i)$ is an interval, which is the convex hull (in $\mathbf{R}$) of a finite collection of the intervals of the Markov partition;
- for each $1 \leq i \leq k$ and n sufficiently big, $\Psi^n(K \cap K_i) = K$ (this means that $\Psi|K$ is topologically mixing).

For a given Cantor set K as above there are Markov partitions; one can even make the intervals K_i as small as one wishes. This follows from the construction of Markov partitions for basic sets [B, 1975]. In the two-dimensional context this construction can be much simplified, see Appendix II.

Here we only indicate how to make such a Markov partition when our basic set Λ is the horseshoe (see Chapter I). In this case $W^u(p)$ and $W^s(p)$ are as indicated below, and $\Lambda = \overline{W^u(p) \cap W^s(p)}$.



In $W^s(p)$ we indicated 16 intersections with $W^u(p)$ (numbered from 0 to 15), they are all in $\alpha(K_0)$. In a "separate copy of $\mathbf{R}$" we indicate the inverse images of these points by α and indicate the intervals $K_1, \ldots, K_8$ of the Markov partition.

In the figure with $W^u(p)$ and $W^s(p)$ we indicated $\alpha(K_0)$ and U, the neighbourhood of Λ on which we assume $\mathcal{F}^u$ to be defined; note that with this choice of U and $\alpha(K_0)$, the projection π (projecting U along fibres of $\mathcal{F}^u$ to $\alpha(K_0)$) is uniquely defined. As expanding map we take

$$\Psi = \alpha^{-1} \circ \pi \circ \varphi^{-1} \circ \alpha.$$

The action of Ψ on the points $0, \ldots, 15$ is then given by:

0	1	2	3	4	5	6	7
↓	↓	↓	↓	↓	↓	↓	↓
0	3	4	7	8	11	12	15
↑	↑	↑	↑	↑	↑	↑	↑
15	14	13	12	11	10	9	8

Taking as intervals of the Markov partition $K_1, \ldots, K_8$ as indicated in the above figure, their images are $\Psi(K_i) = I_i$. From this it is simple to verify that $K_1, \ldots, K_8$ forms indeed a Markov partition.

Observe that if $K_1, \ldots, K_k$ is a Markov partition of a dynamically defined Cantor set K with expanding map Ψ, one gets a Markov partition with more and shorter intervals by just taking as new intervals connected components of $\Psi^{-1}(K_i) \cap K_j$.

So far we have seen the basic properties of dynamically defined Cantor sets. For the purpose of what follows it is convenient to use these properties as definition.

Definition: *A dynamically defined Cantor set* is a Cantor set $K \subset \mathbf{R}$, together with

- a real number λ, the scaling factor, $0 < |\lambda| < 1$, such that λK is a neighbourhood of 0 in K;
- a map $\Psi : K \to K$ having a $C^{1+\epsilon}$ expansive extension to a neighbourhood of K;
- a Markov partition $K_1, \ldots, K_k$.

Observe that although in this chapter we will not make use of the scaling property in the definition, this condition will play a key role in Chapter V.

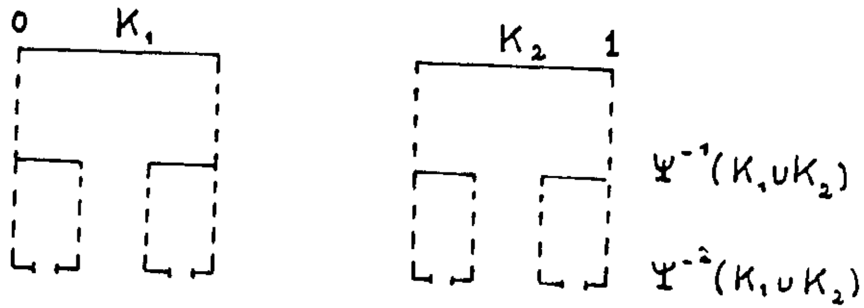
Examples: In each of the examples below we *define* the Cantor set by a Markov partition and expanding map. Observe that we can always associate to a Markov partition $\{K_1, \ldots, K_k\}$ and an expanding map Ψ the Cantor set $K = \bigcap_{i=0}^{\infty} \Psi^{-i}\,(K_1 \cup \cdots \cup K_k)$. Further, we consider only examples where $\Psi|K_i$ is affine, i.e. has constant derivative.

Our first example is the *mid-α-Cantor set*. In this case

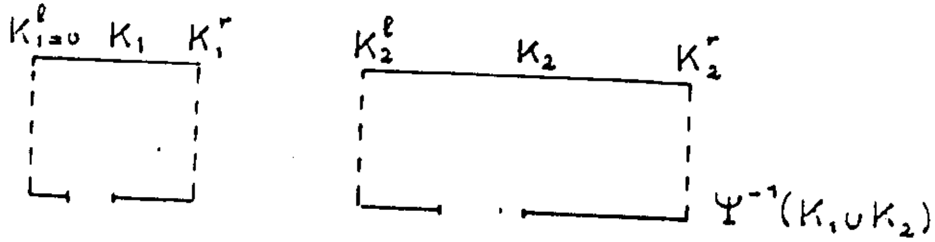
$$K_1 = \left[0, \frac{1}{2}(1-\alpha)\right]$$

$$K_2 = \left[\frac{1}{2}(1+\alpha), 1\right]$$

and $\Psi|K_i$ maps K_i affinely to $[0,1]$; the scaling constant can be taken as $\lambda = \frac{1}{2}(1-\alpha)$. For $\alpha = 1/3$ this is the most well known Cantor set; in any case, for this construction one needs $0 < \alpha < 1$.

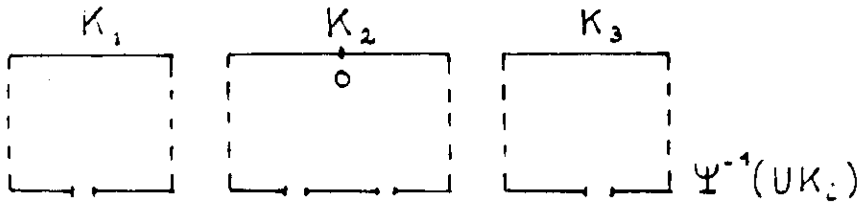


The second example, or rather a class of examples, covers the *affine Cantor sets*. They are defined by a sequence of intervals $K_1, \ldots, K_k$ with endpoints K_i^ℓ, K_i^r so that $0 = K_1^\ell < K_1^r < K_2^\ell < K_2^r < K_3^\ell < \cdots < K_k^r$; also $\Psi|K_i$ maps K_i affinely onto $[0, K_k^r]$; as scaling constant one can take $\lambda = K_1^r / K_k^r$.



Finally we define *generalized affine Cantor sets*. They are obtained as the affine Cantor sets, only now the image $\Psi(K_i)$ may be smaller. If we denote the endpoints of K_i as above by K_i^l and K_i^r with $\cdots < K_i^l < K_i^r < K_{i+1}^l < \cdots$ then $\Psi(K_i)$ should just be an interval of the form $[K_{j_i}^l, K_{j_i'}^r]$ with $j_i \le j_i'$. In this case one still has to verify whether Ψ is expanding, whether there is scaling and whether $\Psi^n(K_i \cap K) = K$ for big n. In the special example below we have

$$\Psi(K_1 \cap K) = (K_1 \cup K_2) \cap K$$
$$\Psi(K_2 \cap K) = K$$
$$\Psi(K_3 \cap K) = (K_2 \cup K_3) \cap K$$



One sees that Ψ is expanding, $\Psi^n(K_i \cap K) = K$ for $n \ge 2$. In order to have scaling one needs 0 to be a fixed point of the affine map $\Psi|K_2$; here we don't require $0 = K_1^l$.

Bounded distortion property: The above examples are special in the sense that Ψ is affine on each K_i. This is in general not the case. However, as we shall see, the distortions due to the fact that the derivatives of the iterates of Ψ are not locally constant, can be bounded in a very strong sense. It is for these estimates that we require Ψ to be $C^{1+\epsilon}$.

Theorem. *Let $K \subset \mathbf{R}$ be a dynamically defined Cantor set with expanding map Ψ. Then, given $\delta > 0$, there is $c(\delta) > 0$ such that for all q, $\tilde q$ and $n \ge 1$ with*

a) $|\Psi^n(q) - \Psi^n(\tilde q)| \le \delta$,

b) the interval $[\Psi^i(q), \Psi^i(\tilde q)]$ contained in the domain of Ψ for all $0 \le i \le n-1$, we have $|\log|(\Psi^n)'(q)| - \log|(\Psi^n)'(\tilde q)|| \le c(\delta)$. Moreover $c(\delta)$ converges to zero when $\delta \to 0$.

Proof: From the fact that Ψ is expanding it follows that, for some $\sigma > 1$, $|\Psi^i(q) - \Psi^i(\tilde q)| \le \delta \cdot \sigma^{i-n}$ for $i \le n$. Since Ψ is $C^{1+\epsilon}$ and Ψ' is bounded away from zero, $\log|\Psi'|$ is C^ϵ. Then

$$\begin{aligned}
|\log|(\Psi^n)'(q)| - \log|(\Psi^n)'(\tilde q)|| &= \left|\sum_{i=0}^{n-1} \log|\Psi'(\Psi^i(q))| - \log|\Psi'(\Psi^i(\tilde q))|\right| \\
&\le \sum_{i=0}^{n-1} C\,|\Psi^i(q) - \Psi^i(\tilde q)|^\epsilon \le \sum_{i=0}^{n-1} C\delta^\epsilon \cdot \sigma^{\epsilon(i-n)} \\
&\le C\delta^\epsilon \frac{\sigma^{-\epsilon}}{1-\sigma^{-\epsilon}},
\end{aligned}$$

for some constant $C > 0$. This proves the theorem by taking $c(\delta) = C\delta^\epsilon \frac{\sigma^{-\epsilon}}{1-\sigma^{-\epsilon}}$. ∎

Let us finish this section with some comments about the bounded distortion property. First, we present a geometric consequence of it. Let V be some small open interval intersecting K. Since Ψ is topologically mixing (see definition of Markov partition), there is $n \ge 1$ such that $\Psi^n(V \cap K) = K$. Take $q_0, q_1, q_2 \in V \cap K$ close enough to each other, so that the intervals $(\Psi^i(q_0), \Psi^i(q_j))$ are contained in the domain of Ψ for $0 \le i \le n-1$ and $j = 1, 2$. By the mean value theorem, there are $q \in (q_0, q_1)$, $\tilde q \in (q_0, q_2)$ such that $|\Psi^n(q_0) - \Psi^n(q_1)| = |q_0 - q_1| \cdot |(\Psi^n)'(q)|$ and $|\Psi^n(q_0) - \Psi^n(q_2)| = |q_0 - q_2| \cdot |(\Psi^n)'(\tilde q)|$. Then, from the theorem above, we get

$$e^{-c}\frac{|q_0 - q_1|}{|q_0 - q_2|} \le \frac{|\Psi^n(q_0) - \Psi^n(q_1)|}{|\Psi^n(q_0) - \Psi^n(q_2)|} \le e^c \frac{|q_0 - q_1|}{|q_0 - q_2|}$$

for some $c > 0$ independent of n, V and the points involved. So Ψ^n essentially preserves ratios of distances between close points: they change but not by more than a uniform multiplicative constant. This means that, up to a bounded distortion, small parts of K are just reproductions of big parts of K in a smaller scale.

Our second remark concerns the differentiability assumptions in the statement and proof of the theorem above. Clearly we used the assumption that Ψ is $C^{1+\epsilon}$; in general if Ψ is only C^1 the theorem is not valid. In order to obtain Ψ to be of class $C^{1+\epsilon}$, it is enough to require our 2-dimensional diffeomorphism φ to be of class C^3 since in this case the tangent lines to the leaves of the foliation, and hence the foliation itself, is of class $C^{1+\epsilon}$. *It is important to observe that Ψ still has the bounded distortion property even when the surface diffeomorphism φ is of class C^2.* The first proof of this fact, which is due to Newhouse and essentially contained in [**N**, 1970; **N**, 1979], follows from the fact that the iterates Ψ^n can be obtained by iterating φ^{-1} n times and *only then* projecting along the leaves of $\mathcal{F}^u$. To illustrate this we consider the basic set Λ to be the horseshoe. Take $\alpha(K_0) \subset W^s(p)$ and U, a neighbourhood of Λ, as in the first figure of the present chapter, and let π denote the projection of U to $\alpha(K_0)$ along the foliation $\mathcal{F}^u$. The fact that this foliation is φ^{-1}-invariant implies that $\pi\varphi^{-1}\pi(x) = \pi\varphi^{-1}(x)$, whenever all these projections are defined. From this, by induction it follows that $(\alpha \circ \Psi \circ \alpha^{-1})^n = \pi \circ \varphi^{-n}$, i.e. $\Psi^n = \alpha^{-1} \circ \pi \circ \varphi^{-n} \circ \alpha$, for all $n \geq 1$ (case $n = 1$ is just the definition of Ψ). *Since, in this expression, the projection π appears only once, its contribution to the distortion of Ψ^n is bounded.* On the other hand the distortion of φ^{-n} can be estimated by the same argument as in the proof of the theorem, since φ^{-1} is, by assumption, C^2 (and so $C^{1+\epsilon}$). In this way one proves that Ψ has the bounded distortion property. Another proof of this fact consists in showing directly that π, restricted to components of $W^s(p) \cap U$, and so Ψ, is of class $C^{1+\epsilon}$, even if φ is only C^2—this was communicated to us by Viana.

§2. Numerical invariants of Cantor sets

In this section we define four numerical invariants for Cantor sets, namely Hausdorff dimension, limit capacity, thichness and denseness. Then we discuss the Lebesgue measure of the difference of two Cantor sets in the real line, in terms of these invariants and finally

we provide some important relations between these invariants when applied to the same Cantor set.

Before we can define Hausdorff dimension, we need to introduce some preliminary notions. Let $K \subset \mathbf{R}$ be a Cantor set and $\mathcal{U} = \{U_i\}_{i\in I}$ a finite covering of K by open intervals in $\mathbf{R}$. We define the diameter $\text{diam}(\mathcal{U})$ of $\mathcal{U}$ as the maximum of $\ell(U_i)$, $i \in I$, where $\ell(U_i)$ denotes the length of U_i. Define $H_\alpha(\mathcal{U}) = \sum_{i\in I} \ell_i^\alpha$. Then the *Hausdorff α-measure* of K is

$$m_\alpha(K) = \lim_{\epsilon\to 0}\left(\inf_{\substack{\mathcal{U} \text{ covers } K\\ \text{diam}(\mathcal{U})<\epsilon}} H_\alpha(\mathcal{U})\right).$$

It is not hard to see that there is a unique number, the *Hausdorff dimension* of K, denoted by $HD(K)$, such that for $\alpha < HD(K)$, $m_\alpha(K) = \infty$ and for $\alpha > HD(K)$, $m_\alpha(K) = 0$.

In order to define the limit capacity, let $N_\epsilon(K)$, K again a Cantor set in $\mathbf{R}$, be the minimal number of intervals of length ϵ needed to cover K. Then the *limit capacity* of K, denoted by $d(K)$, is defined as

$$d(K) = \limsup_{\epsilon\to 0} \frac{\log N_\epsilon(K)}{-\log \epsilon}.$$

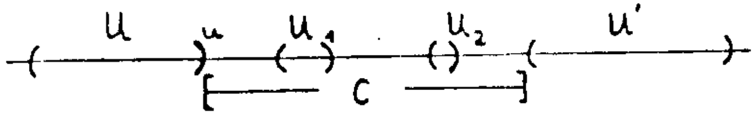
For the mid-α-Cantor set, the first of our examples in the previous paragraph, one can verify that the Hausdorff dimension and limit capacity are both equal to $\log 2/(\log 2 - \log(1-\alpha))$. This will also follow from a more general result to be discussed later.

Both these notions of Hausdorff dimension and limit capacity can be immediately extended to arbitrary subsets of higher dimensional Euclidean spaces (or even metric spaces); in the non-compact case one considers countable (instead of finite) coverings.

To define thickness, we consider the *gaps* of K: a gap of K is a connected component of $\mathbf{R} \setminus K$; a bounded gap is a bounded connected component of $\mathbf{R} \setminus K$. Let U be any bounded gap and u be a boundary point of U, so $u \in K$. Let C be the *bridge* of K at u, i.e. the maximal interval in $\mathbf{R}$ such that:

- u is a boundary point of C;

- C contains no point of a gap U' whose length $\ell(U')$ is at least the length of U.



In the figure U, U', U_1, U_2 are gaps of K, $\ell(U') > \ell(U)$ and C is the bridge of K at u.

The thickness of K at u is defined as $\tau(K,u) = \ell(C)/\ell(U)$. The *thickness* of K, denoted by $\tau(K)$, is the infimum over these $\tau(K,u)$ for all boundary points u of bounded gaps.

Let us provide an equivalent definition of thickness which was actually the one used by Newhouse in [N. 1979]. Incidentally and curiously, M. Hall [H, 1947] had earlier used the same concept in the context of Number Theory.

Define a *presentation* of the Cantor set K as above to be an ordering $\mathcal{U} = \{U_n\}$ of the bounded gaps of K. For $u \in \partial U_n$, let the $\mathcal{U}$-component of K at u be the connected component C of $I - (U_1 \cup \cdots \cup U_n)$ that contains u. Here, I indicates the minimal (closed) interval of $\mathbf{R}$ containing K. For each such u, denote $\tau(K,\mathcal{U},u) = \ell(C)/\ell(U)$. Then, one can check that the thickness of K is given by

$$\tau(K) = \sup_{\mathcal{U}} \inf_{u} \tau(K,\mathcal{U},u),$$

where the infimum is taken over all boundary points of finite gaps of K and the supremum over all presentations of K. Actually the equality follows from the fact that for any presentation $\mathcal{U} = \{U_n\}$, with $\ell(U_n) \leq \ell(U_m)$ for all $n > m$, the supremum in the formula above is assumed.

We define the *denseness* of K, denoted by $\theta(K)$, as

$$\theta(K) = \inf_{\mathcal{U}} \sup_{u} \tau(K,\mathcal{U},u).$$

For any two presentations $\mathcal{U} = \{U_n\}$ and $\mathcal{U}' = \{U'_n\}$ of a Cantor set K, we have $\sup_u \tau(K,\mathcal{U},u) \geq \inf_u \tau(K,\mathcal{U}',u)$ (take for u the boundary points of U_1), and so we always have $\tau(K) \leq \theta(K)$.

We will show latter in the present section that if the thickness is large then the Hausdorff dimension of the Cantor set is also large (i.e., close to one). We will also provide examples showing that the converse is not generally true. If, however, we substitute thickness by denseness then we can state a kind of converse: for a Cantor set in the line, small denseness implies small Hausdorff dimension.

Now we come to the discussion of the Lebesgue measure of the difference of two Cantor sets. Let K_1, K_2 be subsets of $\mathbf{R}$. We define their *difference* as

$$K_1 - K_2 = \{t \in R \mid \exists k_1 \in K_1, k_2 \in K_2, \text{ such that } k_1 - k_2 = t\}$$

Proposition. *Let $K_1, K_2 \subset \mathbf{R}$ be Cantor sets with limit capacity d_1 and d_2. If $d_1 + d_2 < 1$, then the Lebesgue measure of $K_1 - K_2$ is zero.*

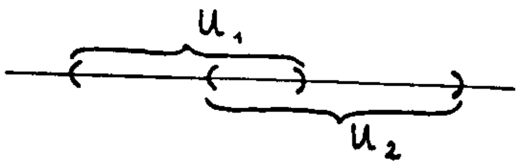
Proof: Let d'_1, d'_2 be numbers such that $d_1 < d'_1$, $d_2 < d'_2$ and $d'_1 + d'_2 < 1$. Then there is an ε_0 such that for $0 < \varepsilon < \varepsilon_0$, K_i can be covered with $\varepsilon^{-d'_i}$ intervals of length ε; this follows directly from the definition of limit capacity. The difference of two intervals of length ε is an interval of length 2ε. So $K_1 - K_2$ is contained in the union of $(\varepsilon^{-d'_1} \cdot \varepsilon^{-d'_2})$ intervals of length 2ε. The total length of these intervals, disregarding overlap, is $2 \cdot \varepsilon^{1-d'_1-d'_2}$. Since $d'_1 + d'_2 < 1$, this can be made arbitrarily small by choosing ε small. Hence the Lebesgue measure of $K_1 - K_2$ is zero. ∎

Gap Lemma. *Let $K_1, K_2 \subset \mathbf{R}$ be Cantor sets with thickness τ_1 and τ_2. If $\tau_1 \cdot \tau_2 > 1$, then one of the following three alternatives occurs: K_1 is contained in a gap of K_2; K_2 is contained in a gap of K_1; $K_1 \cap K_2 \neq \phi$.*

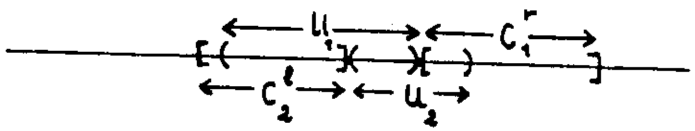
Proof: We assume that neither of the two Cantor sets is contained in a gap of the other and we assume that $K_1 \cap K_2 = \phi$, and derive a contradiction from this. If U_1, U_2 are bounded gaps of K_1, K_2, we call (U_1, U_2) a gappair if U_2 contains exactly one boundary

point of U_1 (and vice-versa); U_1 and U_2 are said to be linked in this case. Since neither of the Cantor sets is contained in a gap of the other and since they are disjoint, there is a gappair. Given such a gappair (U_1, U_2) we construct:
a point in $K_1 \cap K_2$,
or a different gappair (U_1', U_2) with $\ell(U_1') < \ell(U_1)$,
or a different gappair (U_1, U_2') with $\ell(U_2') < \ell(U_2)$.
This leads to a contradiction: even if we don't find a point in $K_1 \cap K_2$ after applying this construction a finite number of times, we get a sequence of gappairs $(U_1^{(i)}, U_2^{(i)})$ such that $\ell(U_1^{(i)})$ or $\ell(U_2^{(i)})$ decreases and hence, since the sum of all the lengths of bounded gaps is finite, it goes to zero. Assuming $\ell(U_1^{(i)})$ goes to zero, take $q_i \in U_1^{(i)}$: any accumulation point of $\{q_i\}$ belongs to $K_1 \cap K_2$.

Now we come to the announced construction. Let the relative position of U_1 and U_2 be as indicated.



Let C_j^ℓ and C_j^r be the bridges of K_j at the boundary points of U_j, $j = 1, 2$. Since $\tau_1 \cdot \tau_2 > 1$, $\frac{\ell(C_1^r)}{\ell(U_1)} \cdot \frac{\ell(C_2^\ell)}{\ell(U_2)} > 1$. So $\ell(C_1^r) > \ell(U_2)$ or $\ell(C_2^\ell) > \ell(U_1)$, or both. Therefore the right endpoint of U_2 is in C_1^r or the left endpoint of U_1 is in C_2^ℓ, or both. Suppose the first. Let u be the right endpoint of U_2. If $u \in K_1$ then we are done, since $u \in K_2$ anyway. If $u \notin K_1$, then u is contained in a gap U_1' of K_1 with $\ell(U_1') < \ell(U_1)$ and (U_1', U_2) is the required gappair. This completes the proof. ∎



Remark: Let now I_1 and I_2 be minimal closed intervals such that $K_1 \subset I_1$ and $K_2 \subset I_2$. We say that K_1 and K_2 are *linked* if I_1 and I_2 are linked. If $\tau(K_1) \cdot \tau(K_2) > 1$ and if K_1 and K_2 are linked, then $K_1 \cap K_2 \neq \phi$ (since neither K_1 can be contained in a gap of K_2 nor K_2 in a gap of K_1). Since being linked is an open condition, it follows that whenever $\tau(K_1) \cdot \tau(K_2) > 1$, then $K_1 - K_2$ has interior points.

Theorem. *Let K_1, K_2 be Cantor sets in $\mathbf{R}$ with Hausdorff dimension h_1, h_2. If $h_1 + h_2 > 1$ then $(K_1 - \lambda K_2)$ has positive Lebesgue measure, for almost every $\lambda \in \mathbf{R}$ (in the Lebesgue measure sense).*

Before going into the proof of the theorem, we first observe that from the assumption on h_1, h_2 it follows that $HD(K_1 \times K_2) \geq HD(K_1) + HD(K_2) > 1$ (see [**F**, 1985]). Also, let us see how we can state this result in a similar but slightly different way. For $\lambda \in \mathbf{R}$ take $\theta \in (-\pi/2, +\pi/2)$ such that $\lambda = -\mathrm{tg}\, \theta$. Let π_θ denote the orthogonal projection of $\mathbf{R}^2$ onto the straight line L_θ which contains $v_\theta = (\cos\theta, \sin\theta)$. If we identify $\mathbf{R}$ with L_θ through $\mathbf{R} \ni x \mapsto x \cdot v_\theta$ then $\pi_\theta(k) = k \cdot v_\theta = \cos\theta \cdot k_1 + \sin\theta \cdot k_2$, for $k = (k_1, k_2) \in \mathbf{R}^2$. By our choice of θ we get $\pi_\theta(K_1 \times K_2) = \cos\theta(K_1 - \lambda K_2)$. Since $\cos\theta \neq 0$ this shows that the theorem above can be rephrased in the following (slightly stronger) form.

Theorem. *Let $K \subset \mathbf{R}^2$ be such that $HD(K) > 1$ and $\pi_\theta : \mathbf{R}^2 \to \mathbf{R}$ be as above. Then $\pi_\theta(K)$ has positive Lebesgue measure for almost every $\theta \in (-\pi/2, +\pi/2)$ (in the Lebesgue measure sense).*

This result was first proved by Marstrand [**M**, 1954]. The argument that we present here, which uses ideas from potential theory, is due to Kaufman and can be found in [**F**, 1985].

Proof: Let $d = HD(K) > 1$. We first assume that $0 < m_d(K) < \infty$ and that for some $C > 0$

$$m_d(K \cap B_r(x)) \leq C r^d \tag{1}$$

for all $x \in \mathbf{R}^2$ and $0 < r \leq 1$. Let μ be the finite measure on $\mathbf{R}^2$ defined by $\mu(A) = m_d(A \cap K)$, for A a measurable subset of $\mathbf{R}^2$. Let us, for $-\pi/2 < \theta < \pi/2$, denote by μ_θ the (unique) measure on $\mathbf{R}$ such that $\int f d\mu_\theta = \int (f\circ\pi_\theta)d\mu$ for every continuous function f. The theorem will follow, if we show that the support of μ_θ has positive Lebesgue measure for almost all $\theta\in(-\pi/2,\pi/2)$, since this support is clearly contained in $\pi_\theta(K)$. To do this we use the following fact.

Lemma. *Let η be a finite measure with compact support on* $\mathbf{R}$ *and* $\hat\eta(p) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}e^{-ixp}d\eta(x)$, *for* $p\in\mathbf{R}$ *($\hat\eta$ is the Fourier transform of η). If* $0<\int_{-\infty}^{+\infty}|\hat\eta(p)|^2\,dp<\infty$ *then the support of η has positive Lebesgue measure.*

Proof of the Lemma: The assumption that $\hat\eta$ is square-integrable implies (Plancherel's theorem) that $\varphi(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}e^{ixp}\hat\eta(p)dp$ is a well-defined, square-integrable function on $\mathbf{R}$ and $d\eta=\varphi dx$. Moreover $\int_{-\infty}^{+\infty}|\varphi(x)|^2\,dx=\int_{-\infty}^{+\infty}|\hat\eta(p)|^2\,dp>0$ and so the support of η, which is equal to the support of φ, can not have Lebesgue measure zero. This proves the lemma. ∎

Returning to the proof of the theorem we now show that, for almost any $\theta\in(-\pi/2,+\pi/2)$, $\hat\mu_\theta$ is square-integrable. From the definitions we have

$$\begin{aligned}|\hat\mu_\theta(p)|^2 &= \frac{1}{2\pi}\iint e^{i(y-x)p}d\mu_\theta(x)d\mu_\theta(y)\\ &= \frac{1}{2\pi}\iint e^{ip(v-u)\cdot v_\theta}d\mu(u)d\mu(v).\end{aligned}$$

Then

$$|\hat\mu_\theta(p)|^2+|\hat\mu_{\theta+\pi}(p)|^2=\frac{1}{\pi}\iint\cos(p(v-u)\cdot v_\theta)d\mu(u)d\mu(v)$$

and so

$$\begin{aligned}\int_0^{2\pi}|\hat\mu_\theta(p)|^2\,d\theta &= \frac{1}{2\pi}\int_0^{2\pi}\iint\cos(p(v-u)\cdot v_\theta)d\mu(u)d\mu(v)d\theta\\ &= \frac{1}{2\pi}\iint\left(\int_0^{2\pi}\cos(p(v-u)\cdot v_\theta)d\theta\right)d\mu(u)d\mu(v).\end{aligned}$$

Note that the integral on θ above does not depend on the direction of $(v-u)$. We now introduce the Bessel function $J_0(z)=\frac{1}{2\pi}\int_0^{2\pi}\cos(z\cos\theta)d\theta$ and write

$$\int_0^{2\pi}|\hat\mu_\theta(p)|^2\,d\theta=\iint J_0(p\,\|v-u\|)d\mu(u)d\mu(v).$$

Integrating on p and using Fubini's theorem we get

$$\begin{aligned}\int_{-a}^{+a}\int_0^{2\pi}|\hat\mu_\theta(p)|^2\,d\theta dp &= \iiint_{-a}^{+a}J_0(p\,\|v-u\|)dpd\mu(u)d\mu(v)\\ &= \iint\left(\int_{-a\|v-u\|}^{+a\|v-u\|}J_0(z)dz\right)\cdot\frac{1}{\|v-u\|}d\mu(u)d\mu(v).\end{aligned}$$

Now, it is well known that $\int_{-\infty}^{+\infty}J_0(z)dz$ is convergent. So, in particular, we can write

$$\int_{-a}^{+a}\int_0^{2\pi}|\hat\mu_\theta(p)|^2\,d\theta\,dp\le A\iint\frac{d\mu(u)d\mu(v)}{\|v-u\|}$$

for some $A>0$ independent of a.

We observe that the integral on the right-hand side is finite. To show this fix $\alpha\in(0,1)$. Then, by condition (1) at the beginning of the proof we have

$$\begin{aligned}\int\frac{d\mu(v)}{\|u-v\|} &= \int_{\|u-v\|\ge1}\frac{d\mu(v)}{\|u-v\|}+\sum_{n=1}^{\infty}\int_{\alpha^n\le\|u-v\|<\alpha^{n-1}}\frac{d\mu(v)}{\|u-v\|}\\ &\le \mu(\mathbf{R}^2)+\sum_{n=1}^{\infty}\alpha^{-n}\mu(B_{\alpha^{n-1}}(u))\\ &\le \mu(\mathbf{R}^2)+\frac{C}{\alpha-\alpha^d},\quad\text{for all}\quad u\in\mathbf{R}^2,\end{aligned}$$

and so

$$\iint\frac{d\mu(u)d\mu(v)}{\|u-v\|}\le\mu(\mathbf{R}^2)\left[\mu(\mathbf{R}^2)+\frac{C}{\alpha-\alpha^d}\right]<\infty.$$

Using Fubini's theorem once more and letting $a\to+\infty$, we get

$$\int_0^{2\pi}\left(\int_{-\infty}^{+\infty}|\hat\mu_\theta(p)|^2\,dp\right)d\theta\le A\iint\frac{d\mu(u)d\mu(v)}{\|v-u\|}<\infty$$

and so $\int_{-\infty}^{+\infty} |\hat{\mu}_\theta(p)|^2\, dp < \infty$ for almost any $\theta \in (-\pi/2, +\pi/2)$. On the other hand we must have $\int_{-\infty}^{+\infty} |\hat{\mu}_\theta(p)|^2\, dp > 0$, for all $\theta \in (-\pi/2, +\pi/2)$. In fact if, for some θ, this integral were zero then (recall the notation in the proof of the lemma) $\int_{-\infty}^{+\infty} |\varphi(x)|^2\, dx = 0$, and so $\varphi = 0$ almost everywhere. Since $d\mu_\theta = \varphi dx$ this would imply $\mu_\theta(\mathbf{R}) = \int_{-\infty}^{+\infty} \varphi(x) dx = 0$ and so $\mu(\mathbf{R}^2) = 0$, which contradicts our assumption that the Hausdorff d-measure of K is positive.

This, according to the lemma, proves the theorem, when K has positive, finite d-measure and satisfies condition (1). We now reduce the general case to this one. Take $1 < d' < d$ (so that $m_{d'}(K) = \infty$) and $K' \subset K$ such that $0 < m_{d'}(K') < \infty$ and condition (1) above is satisfied with K' and d' in place of K and d. Such a K' always exists by Theorem 5.6 in [**F**, 1985]. Then the argument above can be carried out with d and K replaced by d' and K', to conclude that $\pi_\theta(K')$ has positive Lebesgue measure for almost all θ, and so the same holds for $\pi_\theta(K)$. ∎

Remark: It is worthwile to point out that if K_1, K_2 are dynamically defined Cantor sets then the Hausdorff d-measure of $K = K_1 \times K_2$ is positive and finite and m_d satisfies condition (1) above, where $d = HD(K_1) + HD(K_2)$. We will further discuss this fact later in this chapter.

Next we come to the relations between the different invariants when applied to the same Cantor set.

Proposition. *Let $K \subset \mathbf{R}$ be a Cantor set. Then $d(K) \geq HD(K)$.*

Proof: For any $d' > d(K)$ and ε sufficiently small, there is a covering of K with $\varepsilon^{-d'}$ intervals of length ε. For such a covering $\mathcal{U}$, and $d'' > d'$, we have $H_{d''}(\mathcal{U}) = \varepsilon^{-d'} \cdot \varepsilon^{d''}$. For ε going to zero, this last expression goes to zero. This means that for any $d'' > d(K)$, the Hausdorff d'' measure of K is zero, and the proposition follows. ∎

Observe that this same argument proves a somewhat stronger fact: for any Cantor

set $K \subset \mathbf{R}$,

$$HD(K) \leq \liminf_{\varepsilon \to 0} \frac{\log N_\varepsilon(K)}{-\log \varepsilon}.$$

In particular it follows that whenever $HD(K) = d(K)$ then

$$d(K) = \lim_{\varepsilon \to 0} \frac{\log N_\varepsilon(K)}{-\log \varepsilon}$$

(and not just "lim sup"). This is always the case if K is dynamically defined.

Theorem. *(See [**T**, 1988], [**MM**, 1983]). Let $K \subset R$ be a dynamically defined Cantor set. Then $d(K) = HD(K)$.*

Before giving the formal proof, we want to indicate why the theorem is true in the (easier) case of an affine Cantor set; the proof for the general case is based on the same ideas. So let K be the Cantor set defined by the intervals $K_1, \ldots, K_k$ with endpoints $K_1^\ell = 0 < K_1^r < K_2^\ell < \cdots < K_k^r$ and the affine expanding maps $\Psi_i : K_i \to [0, K_k^r]$. We denote the factor, by which distances are multiplied under Ψ_i, by λ_i. So $\lambda_i = \Psi_i' = K_k^r/(K_i^r - K_i^\ell)$. The idea is now to show that both $d(K)$ and $HD(K)$ are equal to the number d for which $\sum \lambda_i^{-d} = 1$. Since $HD(K) \leq d(K)$, we only have to show that $d(K) \leq d$ and $HD(K) \geq d$.

First we indicate why $d(K) \leq d$. Suppose that, for some $\tilde{d} > d$, K can be covered by $\varepsilon^{-\tilde{d}}$ intervals of length ε, whenever $\varepsilon \leq 1$. Then, using the maps Ψ_i^{-1}, we can cover $K \cap K_i$ by $\varepsilon^{-\tilde{d}}$ intervals of length $(\lambda_i^{-1} \cdot \varepsilon)$ for $\varepsilon \leq 1$ or, in other words, by $(\lambda_i^{-\tilde{d}} \cdot \varepsilon^{-\tilde{d}})$ intervals of length ε for $\varepsilon \leq \lambda_i^{-1}$. Since $K = \bigcup (K \cap K_i)$, we can cover K by $((\sum \lambda_i^{-\tilde{d}}) \cdot \varepsilon^{-\tilde{d}})$ intervals of length ε for all $\varepsilon \leq \lambda^{-1}$, where $\lambda = \max \lambda_i$. By induction we find that we can cover K with $((\sum \lambda_i^{-\tilde{d}})^m \cdot \varepsilon^{-\tilde{d}})$ intervals of length ε for all $\varepsilon \leq \lambda^{-m}$.

Since $\tilde{d} > d$, we have $\sum \lambda_i^{-\tilde{d}} < 1$, so for some positive α, $\lambda^{-\alpha} = \sum \lambda_i^{-\tilde{d}}$. Then for $\varepsilon = \lambda^{-m}$ we need no more than

$$\left(\sum \lambda_i^{-\tilde{d}}\right)^m \cdot \varepsilon^{-\tilde{d}} = \varepsilon^{-(\tilde{d}-\alpha)}$$

intervals of length ε to cover K. So $d(K) \leq \tilde{d} - \alpha$. This implies that $d(K) \leq d$.

Next we assume $HD(K) < d$. Then, for $\varepsilon > 0$ and $HD(K) < \hat{d} < d$ we can find a finite covering $\mathcal{U}$ of K (by intervals) such that $H_{\hat{d}}(\mathcal{U}) < \varepsilon$. Taking ε sufficiently small, we may assume that each $U_i \in \mathcal{U}$ intersects only one of the intervals $K_1, \dots, K_k$. The coverings induced by $\mathcal{U}$ on $K \cap K_1, \dots, K \cap K_k$ are denoted by $\mathcal{U}_i$. Clearly $H_{\hat{d}}(\mathcal{U}) = \sum H_{\hat{d}}(\mathcal{U}_i)$. By applying Ψ_i to $\mathcal{U}_i$ we get a covering $\tilde{\mathcal{U}}_i$ of K with $H_{\hat{d}}(\tilde{\mathcal{U}}_i) = \lambda_i^{\hat{d}} \cdot H_{\hat{d}}(\mathcal{U}_i)$. Since $\hat{d} < d$, we have $\sum \lambda_i^{-\hat{d}} > 1$ and, hence, at least one of the $H_{\hat{d}}(\tilde{\mathcal{U}}_i)$ must be smaller than ε. Now $\tilde{\mathcal{U}}_i$ is a new covering of K satisfying again $H_{\hat{d}}(\tilde{\mathcal{U}}_i) < \varepsilon$, but with *less* elements (k, the number of intervals defining K, is at least two and each $\mathcal{U}_i$, $i = 1, \dots, k$, is non-empty). By induction we get such a covering with no elements, which is a contradiction.

Proof: We begin by describing the structure of the proof. Let $\mathcal{R}^1 = \{K_1, \dots, K_m\}$ be a Markov partition for K and, for $n \geq 2$ let $\mathcal{R}^n$ denote the set of connected components of $\Psi^{-(n-1)}(K_i)$, $K_i \in \mathcal{R}^1$. For $R \in \mathcal{R}^n$ take $\lambda_{n,R} = \inf |(\Psi^n)'|_R|$ and $\Lambda_{n,R} = \sup |(\Psi^n)'|_R|$. Define $\alpha_n, \beta_n > 0$ by $\sum_{R \in \mathcal{R}^n} (\Lambda_{n,R})^{-\alpha_n} = C$ and $\sum_{R \in \mathcal{R}^n} (\lambda_{n,R})^{-\beta_n} = 1$ where C is some properly chosen (big) positive number. We show that, for all $n \geq 1$, we have $HD(K) \geq \alpha_n$ and $d(K) \leq \beta_n$. Finally we prove that $(\beta_n - \alpha_n)_n$ converges to zero as $n \to \infty$; this completes the proof of the theorem.

First we fix the constant C. It follows from the definition of β_n that they are uniformly bounded. Let $\hat{\beta} \geq \beta_n$ for all n. Define $C = \sup |(\Psi^k)'|^{\hat{\beta}}$, where k is such that $\Psi^{k+1}(K_i \cap K) = K$ for all $K_i \in \mathcal{R}^1$. Observe that if $\Psi|K_i$ is onto for all i, then we may take $k = 0$ and so $C = 1$.

Now we prove that $d(K) \leq \beta_n$. Let $\beta > d(K)$. Take $\varepsilon_0 > 0$ so that for $0 < \varepsilon \leq \varepsilon_0$ $N_\varepsilon(K) \leq \varepsilon^{-\beta}$, i.e. there is a covering of K by not more than $\varepsilon^{-\beta}$ intervals of length ε. For every $R \in \mathcal{R}^n$, the inverse images by $(\Psi^n|_R)$ of these intervals form a covering of R by intervals of length at most $\varepsilon \lambda_{n,R}^{-1}$. This means that $N_{\varepsilon \lambda_{n,R}^{-1}}(R) \leq \varepsilon^{-\beta}$ for $0 < \varepsilon \leq \varepsilon_0$, or, in other words, $N_\varepsilon(R) \leq \lambda_{n,R}^{-\beta} \cdot \varepsilon^{-\beta}$ for $0 < \varepsilon \leq \lambda_{n,R}^{-1} \cdot \varepsilon_0$. Then $N_\varepsilon(K) \leq \varepsilon^{-\beta} \Big(\sum_{R \in \mathcal{R}^n} \lambda_{n,R}^{-\beta} \Big)$ for

all $0 < \varepsilon \leq \lambda_n^{-1} \varepsilon_0$, where $\lambda_n = \sup_{R \in \mathcal{R}^n} \lambda_{n,R}$. Repeating the argument we get for all $k \geq 1$:

$$N_\varepsilon(K) \leq \varepsilon^{-\beta} \Big(\sum_{R \in \mathcal{R}^n} \lambda_{n,R}^{-\beta} \Big)^k \quad \text{if} \quad 0 < \varepsilon \leq \lambda_n^{-k} \varepsilon_0.$$

This implies

$$d(K) \leq \beta + \lim_{k \to \infty} \frac{\log \Big(\sum_{R \in \mathcal{R}^n} \lambda_{n,R}^{-\beta} \Big)^k}{\log(\lambda_n^k \varepsilon_0^{-1})} = \beta + \frac{\log \Big(\sum_{R \in \mathcal{R}^n} \lambda_{n,R}^{-\beta} \Big)}{\log \lambda_n}$$

and so, making $\beta \to d(K)$,

$$d(K) \leq d(K) + \frac{\log \Big(\sum_{R \in \mathcal{R}^n} \lambda_{n,R}^{-d(K)} \Big)}{\log \lambda_n}.$$

Since $\lambda_n > 1$ this proves that $\sum_{R \in \mathcal{R}^n} \lambda_{n,R}^{-d(K)} \geq 1$, that is, $d(K) \leq \beta_n$.

Now we derive a contradiction from the assumption that $HD(K) < \alpha_n$. Take $HD(K) < \alpha < \alpha_n$. Then there are finite coverings $\mathcal{U}$ of K with arbitrarily small diameter for which $H_\alpha(\mathcal{U})$ is also arbitrarily small. We assume that every element of $\mathcal{U}$ intersects at most one $R \in \mathcal{R}^n$. This will be the case if we require that $H_\alpha(\mathcal{U}) \leq \varepsilon_0$ for some $\varepsilon_0 = \varepsilon_0(n, \alpha) > 0$. We denote $\mathcal{U}_R = \{U \in \mathcal{U} : U \cap R \neq \emptyset\}$. Let as above $k \geq 0$ be such that $\Psi^{k+1}(K_i \cap K) = K$ for all $K_i \in \mathcal{R}^1$. Then, if $H_\alpha(\mathcal{U})$ and hence diam$(\mathcal{U})$ is sufficiently small, $(\Psi^{n+k} \mid R)(\mathcal{U}_R)$ is a well defined covering of K for all $R \in \mathcal{R}^n$. Note that $H_\alpha((\Psi^{n+k}|_R)(\mathcal{U}_R)) \leq (\sup |(\Psi^k)'|)^\alpha \cdot \Lambda_{n,R}^\alpha \cdot H_\alpha(\mathcal{U}_R) \leq C \cdot \Lambda_{n,R}^\alpha \cdot H_\alpha(\mathcal{U}_R)$ (since $\alpha < \alpha_n < \beta_n < \hat{\beta}$). We claim that $H_\alpha((\Psi^{n+k}|_{R_0})(\mathcal{U}_{R_0})) \leq \varepsilon_0$ for some $R_0 \in \mathcal{R}^n$. Otherwise we would have $H_\alpha(\mathcal{U}) = \sum_{R \in \mathcal{R}^n} H_\alpha(\mathcal{U}_R) \geq C^{-1} \sum_{R \in \mathcal{R}^n} \Lambda_{n,R}^{-\alpha} \cdot H_\alpha((\Psi^{n+k}|_R)(\mathcal{U}_R)) \geq C^{-1} \Big(\sum_{R \in \mathcal{R}^n} \Lambda_{n,R}^{-\alpha} \Big) \cdot \varepsilon_0 \geq \Big(C^{-1} \sum_{R \in \mathcal{R}^n} \Lambda_{n,R}^{-\alpha} \Big) \cdot H_\alpha(\mathcal{U})$ which is a contradiction, since, by assumption, $\alpha < \alpha_n$ and so $\sum_{R \in \mathcal{R}^n} \Lambda_{n,R}^{-\alpha} > C$.

In this way we construct, from the initial finite covering $\mathcal{U}$, a new covering $\mathcal{U}' = (\Psi^{n+k}|_{R_c})(\mathcal{U}_{R_c})$ with less elements than $\mathcal{U}$ and such that $H_\alpha(\mathcal{U}') \le \varepsilon_0$. Repeating this argument we eventually obtain a covering of K with no elements at all. This is the required contradiction.

Finally, to prove that $(\beta_n - \alpha_n)_n \to 0$ we first note that, by the bounded distortion property there is $a > 0$, such that $\Lambda_{n,R} \le a \cdot \lambda_{n,R}$, for all $n \ge 1$ and $R \in \mathcal{R}^n$. Take $\delta_n = \dfrac{\alpha_n \log a + \log C}{-\log a + n \log \lambda}$, where $\lambda = \inf|\Psi'| > 1$. Then

$$\begin{aligned}\sum_{R\in\mathcal{R}^n} \lambda_{n,R}^{-(\alpha_n+\delta_n)} &\le a^{(\alpha_n+\delta_n)} \sum_{R\in\mathcal{R}^n} \Lambda_{n,R}^{-\alpha_n}\cdot\Lambda_{n,R}^{-\delta_n}\\ &\le a^{(\alpha_n+\delta_n)}\cdot\lambda^{-n\delta_n}\cdot\sum_{R\in\mathcal{R}^n}\Lambda_{n,R}^{-\alpha_n}\\ &= a^{(\alpha_n+\delta_n)}\cdot\lambda^{-n\delta_n}\cdot C = 1,\end{aligned}$$

by definition of δ_n. It follows that $\beta_n \le \alpha_n + \delta_n$, i.e. $\beta_n - \alpha_n \le \dfrac{\alpha_n \log a + \log C}{n\log\lambda - \log a} \le \dfrac{HD(K)\cdot\log a + \log C}{n\log\lambda - \log a}$. This implies the convergence we have claimed and completes the proof of the theorem. ∎

The above theorem is a consequence of the regularity of dynamically defined Cantor sets. It makes that the propositions on the measure of the difference of two Cantor sets, in terms of limit capacity and Hausdorff dimension, cover, for dynamically defined Cantor sets, almost all cases – the exceptions being $d(K_1)+d(K_2) = 1$ and $K_1 - \lambda K_2$ for exceptional values of λ. Before proceeding with our discussion on the relations between the invariants (dimensions) of a Cantor set, let us explore some consequences of the ideas involved in the proof of this theorem.

First we recall that in the heuristic proof we have the following formula for the Hausdorff dimension and the limit capacity. If K is an affine Cantor set (see the examples in the previous section) with Markov partition $K_1, \ldots, K_k$ and λ_i denotes the (constant) value of $|\Psi'|_{K_i}|$, then $HD(K) = d(K) = d$, d being the unique number such that $\sum \lambda_i^{-d} = 1$. We use this formula to compute the precise value of $HD(K) = d(K)$ in a particular case.

Take K to be an affine Cantor set with Markov partition $K_1, \ldots, K_k$ such that all the K_i have equal length, say $\beta\cdot\operatorname{diam} K$ for $0 < \beta < 1/k$. Since we are assuming that K is affine (and not just generalized affine), Ψ maps each $K_i \cap K$ onto K, so we must have $\lambda_i = \beta^{-1}$ for all i. Therefore $HD(K) = d(K) = \log k/\log\beta^{-1}$. For $k = 2$, since $\beta = (1-\alpha)/2$, we get the formula stated at the beginning of this section. Incidentally, this shows that the dimension of a dynamically defined Cantor set can take any value between 0 and 1. Also, for any $\rho \in (0,1)$, there are diffeomorphisms exhibiting a saddle point p and a basic set Λ with $p \in \Lambda$ such that $HD(\Lambda \cap W^s(p)) = \rho$.

Our second remark concerns the role played by the bounded distortion property. Although we made use of it in the last part of the proof this is not strictly necessary for the theorem above. In fact this result is still true for Cantor sets defined by expanding maps which are only C^1 (and so may not have this property), see [**T**, 1988]. Even more so, if φ is just a C^1 diffeomorphism, still the Cantor sets $W^s(p) \cap \Lambda$ induced by it have their Hausdorff dimension equal to the limit capacity, see [**PV**, 1988]. However, by using the bounded distortion property one can give better estimates for the velocity of convergence of α_n and β_n, than would hold if Ψ were just C^1.

Recall that in the proof of the theorem we showed that, for some $b > 0$,

$$\beta_n - \frac{b}{n} \le \alpha_n \le d(K) = HD(K) \le \beta_n \le \alpha_n + \frac{b}{n}, \text{ for all } n \ge 1. \tag{1}$$

We want to explore some important consequences of this estimate. First observe that, denoting $A = \sup|\Psi'|$ and $d = HD(K) = d(K)$,

$$\sum_{R\in\mathcal{R}^n} \lambda_{n,R}^{-d} = \sum_{R\in\mathcal{R}^n} \lambda_{n,R}^{-\beta_n}\cdot\lambda_{n,R}^{\beta_n - d} \le \sum_{R\in\mathcal{R}^n} \lambda_{n,R}^{-\beta_n}\cdot A^{n(\beta_n - d)}$$

and so

$$\sum_{R\in\mathcal{R}^n} \lambda_{n,R}^{-d} \le A^b < \infty, \text{ for all } n \ge 1. \tag{2}$$

In a similar way,

$$\sum_{R\in\mathcal{R}^n} \Lambda_{n,R}^{-d} \ge CA^{-b} > 0, \text{ for all } n \ge 1. \tag{3}$$

Using these facts we prove

Proposition. *Let $K \subset \mathbf{R}$ be a dynamically defined Cantor set and let $d = HD(K)$. Then $0 < m_d(K) < \infty$. Moreover, there is $c > 0$ such that, for all $x \in K$ and $0 < r \le 1$,*

$$c^{-1} \le \frac{m_d(B_r(x) \cap K)}{r^d} \le c. \tag{4}$$

We point out that the bounded distortion property is fundamental here: contrary to the theorem above, this last proposition wouldn't hold in general if Ψ were only C^1.

Proof: We keep the notations from the proof of the above theorem. Observe that by the mean value theorem and condition (2),

$$H_d(\mathcal{R}^n) = \sum_{R \in \mathcal{R}^n} |\ell(R)|^d \le \sum_{R \in \mathcal{R}^n} \left(\lambda_{n,R}^{-1} \cdot \ell(K)\right)^d \le A^b(\ell(K))^d,$$

where $\ell(K)$ denotes the diameter of K. Since $\operatorname{diam}(\mathcal{R}^n) \to 0$ as $n \to \infty$, this proves that

$$m_d(K) \le A^b(\ell(K))^d < \infty$$

Proving that $m_d(K)$ is positive requires a little more effort. First, we claim that for some $a_1 > 0$ we have

$$(\ell(U))^d \ge a_1 \cdot \sum_{\substack{R \in \mathcal{R}^n \\ R \cap U \neq \phi}} \Lambda_{n,R}^{-d} \tag{5}$$

for every interval U intersecting K and $n \ge 1$ sufficiently large depending on U. To show this we fix $\alpha > 0$ such that the α-neighbourhood of K is contained in the domain of Ψ. Take $k = k(U) \ge 0$ minimal such that

$$\ell(\Psi^k(U)) \ge \alpha.$$

Let $n > k$. Then $S \in \mathcal{R}^{n-k}$ intersects $\Psi^k(U)$ if and only if $S = \Psi^k(R)$ for some $\mathcal{R} \in \mathcal{R}^n$ intersecting U. Moreover, in such case we have

$$\Lambda_{n,R} = \sup |(\Psi^n)'|_R| \ge \inf |(\Psi^k)'|_R| \cdot \sup |(\Psi^{n-k})'|_S|$$
$$\ge \inf |(\Psi^k)'|_{U \cup R}| \cdot \Lambda_{n-k,S}$$

On the other hand, by the mean value theorem we have

$$\ell(\Psi^k(U)) \le \sup |(\Psi^k)'|_U| \cdot \ell(U) \le \sup |(\Psi^k)'|_{U \cup R}| \cdot \ell(U).$$

Observe that, by construction, $\Psi^j(U \cup R)$ is contained in the domain of Ψ for all $0 \le j \le k-1$. The bounded distortion property implies that

$$\sup |(\Psi^k)'|_{U \cup R}| \le a \cdot \inf |(\Psi^k)'|_{U \cup R}|,$$

where a is some positive number independent of U, R and k. From all this and the fact that $d \ge \alpha_{n-k}$, we obtain

$$(\ell(U))^d \ge C^{-1} \sum_{\substack{S \in \mathcal{R}^{n-k} \\ S \cap \Psi^k(U) \neq \phi}} \Lambda_{n-k,S}^{-d} \cdot \ell(U)^d$$
$$\ge C^{-1} \sum_{\substack{R \in \mathcal{R}^n \\ R \cap U \neq \phi}} \Lambda_{n-k,\Psi^k(R)}^{-d} \cdot (\ell(\Psi^k(U)) / \sup |(\Psi^k)'|_{U \cup R}|)^d$$
$$\ge C^{-1} \alpha^d \sum_{\substack{R \in \mathcal{R}^n \\ R \cap U \neq \phi}} \Lambda_{n-k,\Psi^k(R)}^{-d} \cdot a^{-d} \cdot (\inf |(\Psi^k)'|_{U \cup R}|)^{-d}$$
$$\ge C^{-1} \alpha^d a^{-d} \sum_{\substack{R \in \mathcal{R}^n \\ R \cap U \neq \phi}} \Lambda_{n,R}^{-d}.$$

This proves the claim with $a_1 = C^{-1} \alpha^d a^{-d}$.

Let now $\mathcal{U}$ be any finite covering of K. Take $n \ge 1$ such that (5) holds for all $U \in \mathcal{U}$. Then, by (3),

$$H_d(\mathcal{U}) = \sum_{U \in \mathcal{U}} (\ell(U))^d \ge \sum_{U \in \mathcal{U}} a_1 \Big(\sum_{\substack{R \in \mathcal{R}^n \\ R \cap U \neq \phi}} \Lambda_{n,R}^{-d} \Big)$$
$$\ge a_1 \sum_{R \in \mathcal{R}^n} \Lambda_{n,R}^{-d} \ge a_1 C A^{-b}.$$

Since $\mathcal{U}$ is arbitrary, this proves

$$m_d(K) \ge a_1 C A^{-b} > 0.$$

Now we deal with the second part of the proposition. To make the argument more transparent we first derive an estimate for the d-measure of the intervals $R \in \mathcal{R}^n$. For some $a_2 > 1$, depending only on K and Ψ, we have

$$a_2^{-1} \leq \frac{m_d(R \cap K)}{(\ell(R))^d} \leq a_2 \tag{6}$$

for all $R \in \mathcal{R}^n$ and $n \geq 1$. To show this we observe that Ψ^{n-1} maps R diffeomorphically onto some $K_i \in \mathcal{R}^1$. From the definition of Hausdorff measure, we have

$$\lambda^d_{n-1,R} \cdot m_d(R \cap K) \leq m_d(K_i \cap K) \leq \Lambda^d_{n-1,R} \cdot m_d(R \cap K).$$

On the other hand, by the mean value theorem, we have

$$\lambda_{n-1,R} \cdot \ell(R) \leq \ell(K_i) \leq \Lambda_{n-1,R} \cdot \ell(R).$$

Finally, by the bounded distortion property, it follows that

$$\Lambda_{n-1,R} \leq a \cdot \lambda_{n-1,R}$$

with $a > 0$ as above depending only on K and Ψ. From all this we get

$$a^{-d} \cdot \frac{m_d(K_i \cap K)}{(\ell(K_i))^d} \leq \frac{m_d(R \cap K)}{(\ell(R))^d} \leq \frac{m_d(K_i \cap K)}{(\ell(K_i))^d} \cdot a^d.$$

Clearly, $\ell(K_i)$ can be uniformly bounded from zero and infinity, so to prove (6) we only need to show that the same holds for $m_d(K_i \cap K)$. The upper bound is trivial since $m_d(K_i \cap K) \leq m_d(K) < \infty$.

The lower bound follows easily from the fact that, for some $k \geq 0$, $\Psi^{k+1}(K_i \cap K) = K$ and so, again by the definition of Hausdorff measure,

$$m_d(K_i \cap K) \geq \left(\sup\left|(\Psi^{k+1})'\right|\right)^{-d} m_d(K) > 0.$$

Now we prove (4). For $x \in K$ and $0 < r \leq 1$, we let $q = q(x, r) \geq 0$ be minimal such that

$$\Psi^q(B_r(x)) \not\subset B_\alpha(\Psi^q(x))$$

where, as before, $\alpha > 0$ is such that the domain of Ψ contains the α-neighbourhood of K. Then, arguing as above with $B_r(x)$ and Ψ^q in the place of R and Ψ^{n-1}, respectively, we obtain

$$a^{-d} \cdot \frac{m_d(\Psi^q(B_r(x)) \cap K)}{\ell(\Psi^q(B_r(x)))} \leq \frac{m_d(B_r(x) \cap K)}{(2r)^d} \leq \frac{m_d(\Psi^q(B_r(x)) \cap K)}{\ell(\Psi^q(B_r(x)))} \cdot a^d.$$

Again, $\ell(\Psi^q(B_r(x)))$ can be easily bounded: by construction

$$\alpha \leq \ell(\Psi^q(B_r(x))) \leq A \cdot \alpha.$$

Since we also have

$$m_d(\Psi^q(B_r(x)) \cap K) \leq m_d(K) < \infty,$$

it is enough to provide a uniform lower bound for the d-measure of $\Psi^q(B_r(x)) \cap K$. To do this, we observe that by the mean value theorem we have

$$B_{r\Lambda_q}(\Psi^q(x)) \supset \Psi^q(B_r(x)) \supset B_{r\lambda_q}(\Psi^q(x)),$$

where $\lambda_q = \inf\left|(\Psi^q)'|_{B_r(x)}\right|$ and $\Lambda_q = \sup\left|(\Psi^q)'|_{B_r(x)}\right|$. Then, by the definition of q,

$$r\Lambda_q \geq \alpha$$

and so, using the bounded distortion property once again, $\Psi^q(B_r(x)) \supset B_{\alpha a^{-1}}(\Psi^q(x))$. Fix $p \geq 1$ such that $\ell(R) < \alpha a^{-1}$ for all $R \in \mathcal{R}^p$. Then, $\Psi^q(B_r(x))$ must contain the interval $R_0 \in \mathcal{R}^p$ that contains $\Psi^q(x)$. Finally, one proves, as we did before for $K_i \in \mathcal{R}^1$, that $m_d(R \cap K) > 0$ for all $R \in \mathcal{R}^p$. It follows that

$$m_d(\Psi^q(B_r(x)) \cap K) \geq m_d(R_0 \cap K) \geq \inf\{m_d(R \cap K) \mid R \in \mathcal{R}^p\} > 0$$

and this completes the proof of the proposition. ∎

Finally, we prove a two-dimensional version of this proposition, which had been commented following the proof of Marstrand's theorem relating Hausdorff dimension and measure of the difference set. It applies to hyperbolic basic sets of diffeomorphisms on surfaces.

Proposition. *Let K_1, K_2 be dynamically defined Cantor sets and let $d_1 = HD(K_1)$, $d_2 = HD(K_2)$, $d = d_1 + d_2$ and $K = K_1 \times K_2$ in $\mathbf{R}^2$. Then, for some $c > 0$,*

(a) $0 < m_d(K) < \infty$

(b) $c^{-1} \le \frac{m_d(K \cap B_r(x))}{r^d} \le c$ *for all $x \in K$ and $0 < r \le 1$.*

Proof: Take μ to be the product measure $\mu = m_{d_1} \times m_{d_2}$ on K. Clearly, (a) and (b) hold if we replace there m_d by μ. Therefore it is now sufficient to show that μ is equivalent to m_d in the sense that for all Borel subsets $A \subset K$, $\mu(A)/m_d(A)$ is bounded away from zero and infinity. We consider Markov partitions $\mathcal{R}_1, \mathcal{R}_2$ for K_1, K_2 respectively and denote by $\mathcal{R}_i^n$, $i = 1, 2$, the family of connected components of $\Psi_i^{-(n-1)}(L_j)$, $L_j \in \mathcal{R}_i$. We may restrict ourselves to Borel sets of the form

$$A = R_1 \times R_2, \quad R_1 \in \mathcal{R}_1^n, \quad R_2 \in \mathcal{R}_2^n$$

since these sets generate the Borel σ-algebra of K. Let $\mathcal{U} = \{U_{1,j} \times U_{2,j} : 1 \le j \le m\}$ be any finite covering of $A = R_1 \times R_2$ by cubes. Fix $x_{i,j} \in U_{i,j} \cap R_i$, $i = 1, 2$, $1 \le j \le m$ (obviously, we may assume $U_{i,j} \cap R_i \ne \phi$). Then

$$\begin{aligned}\mu(U_{1,j} \times U_{2,j}) &= m_{d_1}(U_{1,j}) \times m_{d_2}(U_{2,j}) \\ &\le m_{d_1}(B_{1,j}) \times m_{d_2}(B_{2,j}) \\ &\le c_1 c_2 \cdot (\ell(U_{1,j}))^{d_1} \cdot (\ell(U_{2,j}))^{d_2}\end{aligned}$$

where $B_{i,j}$ denotes the ball in K_i centered in $x_{i,j}$ and with radius $\ell(U_{i,j})$. Therefore

$$\begin{aligned}\sum_{j=1}^{m} (\mathrm{diam}(U_{1,j} \times U_{2,j}))^d &\ge \sum_{j=1}^{m} (\ell(U_{1,j}))^{d_1} \times (\ell(U_{2,j}))^{d_2} \\ &\ge (c_1 c_2)^{-1} \sum_j \mu(U_{1,j} \times U_{2,j}) \\ &\ge (c_1 c_2)^{-1} \mu(A).\end{aligned}$$

Since $\mathcal{U}$ is arbitrary this proves

$$m_d(A) \ge (c_1 c_2)^{-1} \mu(A).$$

To obtain an inequality in the opposite direction we construct coverings $\mathcal{U}_m$ of $A = R_1 \times R_2$, $m >> n$, as follows. Fix $U_1 \in \mathcal{R}_1^m$, U_1 contained in R_1. For each $x_2 \in R_2$, take $m(U_1, x_2)$ maximal such that if $U_2(U_1, x_2)$ denotes the element of $\mathcal{R}_2^{m(U_1, x_2)}$ containing x_2, then $\ell(U_2(U_1, x_2)) \ge \ell(U_1)$. Clearly, $\{U_2(U_1, x_2) : x_2 \in R_2\}$ contains a finite covering of R_2 by disjoint intervals. Since these $U_2(U_1, x_2)$ are elements of Markov partitions $\mathcal{R}_2^j$, $j \ge 1$, two of then are either disjoint or one is contained in the other. Thus, we can extract a finite subcovering by disjoint elements. We now define $\mathcal{U}_m$ to be the family of sets $U_1 \times U_2(U_1, x_2)$ obtained in this way for all $U_1 \in \mathcal{R}_1^m$ contained in R_1. This is a covering of $R_1 \times R_2$ by disjoint cubes. Moreover, it is not difficult to deduce from the bounded distortion property that there is $0 < b < 1$ (depending only on K_2 and Ψ_2) such that, denoting by $U_2'(U_1, x_2)$ the element of $\mathcal{R}_2^{m(U_1, x_2)+1}$ that contains x_2, we have $\ell(U_2'(U_1, x_2)) \ge b\ell(U_2(U_1, x_2))$. By definition of $U_2'(U_1, x_2)$, we also have $\ell(U_2'(U_1, x_2)) \le \ell(U_1)$. Therefore,

$$\ell(U_1) \ge b \cdot \ell(U_2(U_1, x_2)). \tag{7}$$

Then, by (6) and (7) we get

$$\begin{aligned}\sum_{\mathcal{U}_m} (\mathrm{diam}(U_1 \times U_2(U_1, x_2)))^d &= \sum_{\mathcal{U}_m} \ell(U_2(U_1, x_2))^d \\ &\le b^{-d_1} \sum_{\mathcal{U}_m} \ell(U_1)^{d_1} \ell(U_2(U_1, x_2))^{d_2} \\ &\le b^{-d_1} \sum_{\mathcal{U}_m} (a_2' m_{d_1}(U_1 \cap K_1))(a_2'' m_{d_2}(U_2(U_1, x_2)) \cap K_2)) \\ &= b^{-d_1} a_2' a_2'' \mu(A).\end{aligned}$$

Since the diameter of $\mathcal{U}_m$ may be taken arbitrarily small (by taking m large), we have $m_d(A) \le b^{-d_1} a_2' a_2'' \mu(A)$ and so our argument is complete. ∎

Note that $0 < m_d(K) < \infty$ ((a) in the proposition) is related to the fact that since K_1 and K_2 are dynamically defined, $HD(K_1 \times K_2) = HD(K_1) + HD(K_2)$. This also follows from the previous theorem stating that $d(K_i) = HD(K_i)$, $i = 1, 2$, and the general product formulas $HD(K_1 \times K_2) \ge HD(K_1) + HD(K_2)$ and $d(K_1 \times K_2) \le d(K_1) + d(K_2)$ together with the inequality $d(K) \ge HD(K)$.

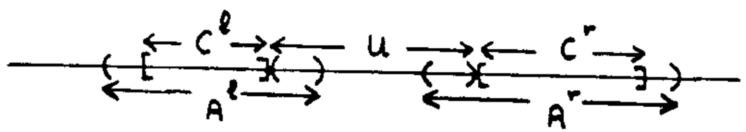
We now establish an interesting relation between Hausdorff dimension and thickness for Cantor sets in the line. In particular, if the thickness is large then the Hausdorff dimension is close to one.

Proposition. *If $K \subset \mathbf{R}$ is a Cantor with thickness τ then $HD(K) \geq (\log 2/\log(2+1/\tau))$.*

Proof: Let $\beta = (\log 2/\log(2+1/\tau))$. We show that $H_\beta(\mathcal{U}) \geq (\operatorname{diam} K)^\beta$ for every finite open covering $\mathcal{U}$ of K, which clearly implies the proposition. The key ingredient in this proof is the following elementary fact:

$$\min\{x^\beta + z^\beta : x \geq 0, z \geq 0, x+z \leq 1, x \geq \tau(1-x-z), z \geq \tau(1-x-z)\} = 1. \qquad (*)$$

We assume from now on that $\mathcal{U}$ is a covering with disjoint intervals. This is no restriction because whenever two elements of $\mathcal{U}$ have non-empty intersection we can replace them by their union, getting in this way a new covering $\mathcal{V}$ such that $H_\beta(\mathcal{V}) \leq H_\beta(\mathcal{U})$. Note that, since $\mathcal{U}$ is an open covering of K, it covers all but a finite number of gaps of K. Let U, a gap of K, have minimal length among the gaps of K which are not covered by $\mathcal{U}$. Let C^ℓ and C^r be the bridges of K at the boundary points of U.



By construction there are A^ℓ, $A^r \in \mathcal{U}$ such that $C^\ell \subset A^\ell$ and $C^r \subset A^r$. Take the convex hull A of $A^\ell \cup A^r$. Then

$$\ell(A^\ell) \geq \ell(C^\ell) \geq \tau \cdot \ell(U) \geq \tau(\ell(A) - \ell(A^\ell) - \ell(A^r))$$

and

$$\ell(A^r) \geq \ell(C^r) \geq \tau \cdot \ell(U) \geq \tau(\ell(A) - \ell(A^\ell) - \ell(A^r))$$

and so, by (*), $(\ell(A^\ell))^\beta + (\ell(A^r))^\beta \geq (\ell(A))^\beta$. This means that the covering $\mathcal{U}_1$ of K obtained by replacing A^ℓ and A^r by A in $\mathcal{U}$ is such that $H_\beta(\mathcal{U}_1) \leq H_\beta(\mathcal{U})$. Repeating the argument we eventually construct $\mathcal{U}_k$, a covering of the convex hull of K with $H_\beta(\mathcal{U}_k) \leq H_\beta(\mathcal{U})$. Since we must have $H_\beta(\mathcal{U}_k) \geq (\dim K)^\beta$, this ends the proof. ∎

Note that in general there can be no nontrivial upper estimates for the Hausdorff dimension in terms of the thickness, even in the dynamically defined case. To see this, recall the earlier example of an affine Cantor set K with Markov partition $\{K_1, \ldots, K_k\}$ all of length $\beta \cdot \operatorname{diam} K$, $0 < \beta < \frac{1}{k}$ and gaps between K_i and K_{i+1} all of length $(1-\beta\cdot k)\cdot \operatorname{diam}(K)/(k-1)$. As we saw, we have $HD(K) = \log k/\log \beta^{-1}$. The thickness can easily be shown to be $\tau(K) = \beta(k-1)/(1-\beta k)$. Now consider a sequence of such Cantor sets characterized by β_k and k such that $\lim_{k\to\infty} k\cdot\beta_k = \alpha \in (0,1)$. Then, as $k \to \infty$, the Hausdorff dimension tends to one while the thickness converges to $\alpha/(1-\alpha)$.

This fact is not really surprising since the thickness was defined as an infimum and so having $\tau(K)$ small gives very little information concerning the Cantor set. As mentioned before, this was our main motivation for introducing a variation of the thickness which we called denseness. We shall prove that Cantor sets with small denseness have small Hausdorff dimension.

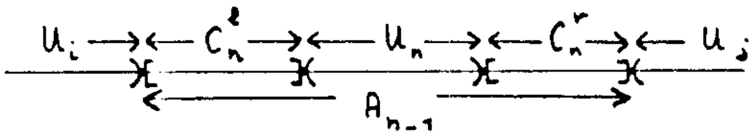
Let us first observe that if K is an affine Cantor set as above and $2\ell - 1 \leq k \leq 2\ell$, then $\theta(K) = (\ell - 1) + \ell\beta(k-1)/(1-\beta k)$. This follows from the fact that the infimun in the definition of $\theta(K)$ is attained on the presentations $\mathcal{U} = \{U_n\}$ of K satisfying

- if $\ell(U_m) > \ell(U_n)$ then $m < n$;
- for U_m, U_n with $\ell(U_m) = \ell(U_n)$ we consider the maximal interval C, if it exists, containing U_m and U_n but containing no points of gaps longer than U_m or U_n; if dist $(U_m, \partial C) >$ dist $(U_n, \partial C)$ then we require $m < n$.

In particular, for such Cantor sets the thickness and the denseness coincide only when $k = 2$.

Proposition. *Let $K \subset \mathbf{R}$ be a Cantor set with denseness θ. Then $HD(K) \leq \log 2/\log(2 + 1/\theta)$.*

Proof: Given $\theta_1 > \theta$, let $\mathcal{U} = (U_n)$ be a presentation of K such that $\sup_u \tau(K, \mathcal{U}, u) \leq \theta_1$. For $n \geq 1$, let $\mathcal{A}_n$ be the covering of K formed by connected components of $I - (U_1 \cup U_2 \cup \cdots \cup U_n)$, where I is the minimal interval containing K. We claim that for $\beta = \log 2/\log(2 + 1/\theta_1)$ the sequence $(H_\beta(\mathcal{A}_n))_n$ is bounded. To show this, we first observe that the only difference between consecutive coverings $\mathcal{A}_{n-1}$ and $\mathcal{A}_n$ is that the interval $A_{n-1} \in \mathcal{A}_{n-1}$ containing U_n is replaced by two new intervals $C_n^\ell, C_n^r \in \mathcal{A}_n$.



Then, we have $H_\beta(\mathcal{A}_n) - H_\beta(\mathcal{A}_{n-1}) = [\ell(C_n^\ell)]^\beta + [\ell(C_n^r)]^\beta - [\ell(A_{n-1})]^\beta$. On the other hand, the assumptions in the proposition imply $\ell(C_n^\ell) \leq \theta_1 \cdot \ell(U_n)$ and $\ell(C_n^r) \leq \theta_1 \cdot \ell(U_n)$. Now using

$$\max\{x^\beta + z^\beta : x \geq 0, z \geq 0, x + z \leq 1, x \leq \theta_1 \cdot (1 - x - z), z \leq \theta_1 \cdot (1 - x - z)\} = 1,$$

we get $[\ell(C_n^\ell)]^\beta + [\ell(C_n^r)]^\beta \leq [\ell(A_{n-1})]^\beta$, that is $H_\beta(\mathcal{A}_n) \leq H_\beta(\mathcal{A}_{n-1})$. Therefore, the sequence $(H_\beta(\mathcal{A}_n))_n$ is nonincreasing and so it is bounded as claimed. Since the diameters of $\mathcal{A}_n$ clearly converge to zero this implies $m_\beta(K) < \infty$ and so $HD(K) \leq \beta = \log 2/\log(2 + 1/\theta_1)$. Since $\theta_1 > \theta$ is arbitrary, the proposition follows. ∎

Proposition. *If K is a dynamically defined Cantor set then $0 < \tau(K) \leq \theta(K) < \infty$ and so $0 < d(K) = HD(K) < 1$.*

Proof: We have already seen that the denseness is always larger than or equal to the thickness. Let us show that for some presentation $\mathcal{U}$ of K

$$0 < \inf_u \tau(K, \mathcal{U}, u) \leq \sup_u \tau(K, \mathcal{U}, u) < \infty. \tag{1}$$

This immediately implies the first statement in the proposition. The second statement is a direct consequence of the first one and the two last propositions above. To construct $\mathcal{U}$ we proceed as follows. Let $\mathcal{R} = \{K_1, \ldots, K_k\}$ be a Markov partition for K and $\tilde{U}_1, \ldots, \tilde{U}_{k-1}$ be the gaps of K between the intervals in this partition. For any gap U of K let $s(U) \geq 0$ be the smallest integer such that $\Psi^{s(U)}(U)$ is not contained in any $K_s \in \mathcal{R}$ ($s(U) = 0$ if and only if $U \in \{\tilde{U}_1, \ldots, \tilde{U}_{k-1}\}$ or U is unbounded). Clearly, given any $\bar{s} \geq 0$ the set of gaps U of K such that $s(U) \leq \bar{s}$ is finite. Therefore we may take $\mathcal{U} = \{U_n\}$ an ordering of the bounded gaps of K such that

$$i \leq j \Rightarrow s(U_i) \leq s(U_j).$$

We now prove that such $\mathcal{U}$ satisfies (1). Let $u \in \partial U_i$ and C be the $\mathcal{U}$-component of K at u. Observe that $\Psi^n(C)$ is contained in some $K_{i_n} \in \mathcal{R}$, for all $0 \leq n \leq s(U_i) - 1$. Otherwise C would contain a gap U_j with $s(U_j) < s(U_i)$ and so $j < i$, which contradicts the definition of $\mathcal{U}$-component. Then by the bounded distortion property there is $a > 0$, depending only on (K, Ψ), such that

$$a^{-1} \cdot \frac{\ell(C)}{\ell(U_i)} \leq \frac{\ell(\Psi^{s(U_i)}(C))}{\ell(\Psi^{s(U_i)}(U_i))} \leq a \cdot \frac{\ell(C)}{\ell(U_i)}$$

i.e.

$$a^{-1} \cdot \frac{\ell(\Psi^{s(U_i)}(C))}{\ell(\Psi^{s(U_i)}(U_i))} \leq \tau(K, \mathcal{U}, u) \leq a \cdot \frac{\ell(\Psi^{s(U_i)}(C))}{\ell(\Psi^{s(U_i)}(U_i))}.$$

To complete the proof it is now sufficient to show that the values of $\dfrac{\ell(\Psi^{s(U_i)}(C))}{\ell(\Psi^{s(U_i)}(U_i))}$ that we obtain in this way can be bounded from zero and ∞. To see this observe first that we must have $\Psi^{s(U_i)}(U_i) \in \{\tilde{U}_1, \ldots, \tilde{U}_{k-1}\}$ and so $\ell(\Psi^{s(U_i)}(U_i))$ can take only a finite set of values. As to $\Psi^{s(U_i)}(C)$, note that its length cannot exceed $\operatorname{diam}(K)$ and that, on the other hand, it must contain some $K_r \in \mathcal{R}$. This last affirmative is proved as follows. Let v be the other boundary point of C and U the gap of K such that $v \in \partial U$. Then either U is unbounded or $U = U_j$ for some $j < i$. In either case we have $s(U) \leq s(U_i)$. It follows that $\Psi^{s(U_i)}(v)$ must be in the boundary of some K_r, which then must be contained in $\Psi^{s(U_i)}(C)$. We conclude that $\ell(\Psi^{s(U_i)}(C))$ is also bounded away from zero and infinity. This completes the proof of the proposition. ∎

It follows easily from the definition of Hausdorff dimension that, for any subset K of $\mathbf{R}$, we have

$$HD(K) < 1 \Longrightarrow m(K) = 0,$$

where m denotes the Lebesgue measure. ***As an important particular case, the dynamically defined Cantor sets have always Lebesgue measure zero.*** This conclusion remains valid for Cantor sets defined by 2-dimensional diffeomorphisms which are *only of class C^2*, since, as observed in the last remark of Section 1, such Cantor sets still have the bounded distortion property.

The situation is rather different for diffeomorphisms which are only once differentiable. We describe briefly a construction, due to Bowen [**B**,1975], of a C^1 diffeomorphism with an invariant horseshoe $\Lambda = \overline{W^s(p) \cap W^u(p)}$, p a hyperbolic fixed point, such that $m(\Lambda) > 0$ and $m(W^s(p) \cap \Lambda) > 0$. In this particular example, $W^s(p) \cap \Lambda$ is invariant under an expanding C^1 map but the bounded distortion property no longer holds.

Given a sequence $\{\beta_n\}$ of positive real numbers satisfying

$$\sum_{n \geq 0} \beta_n < 2 \quad \text{and} \quad \frac{\beta_n}{\beta_{n+1}} \underset{n\to\infty}{\longrightarrow} 1,$$

we construct in $J = [-1, 1]$, by the standard procedure, a Cantor set K_J in such way that at the n^{th}-step we remove 2^n intervals, $J_{n,k}$, $k \in \{1, \ldots, 2^n\}$, of length $\frac{\beta_n}{2^n}$. It is then clear that $m(K_J) = 2 - \sum_{n \geq 0} \beta_n$ is positive. For each $n \geq 1$ and $k \in \{2^{n-1}+1, \ldots, 2^n\}$, define

$$g_{n,k}: J_{n,k} \longrightarrow J_{n-1, k-2^{n-1}}$$

as follows:

(i) $g_{n,k}$ is a C^1 orientation preserving homeomorphism;

(ii) $g'_{n,k}(a_{n,k}) = g'_{n,k}(b_{n,k}) = 2$, where $J_{n,k} = [a_{n,k}, b_{n,k}]$;

(iii) $\sup_{x \in J_{n,k}} |2 - g'_{n,k}(x)| \underset{n\to+\infty}{\longrightarrow} 0$

The choice of the β_n's guarantees that (i)–(iii) coexist and makes possible this construction.

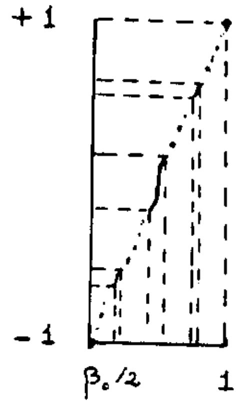
Now, from the above conditions, we can continuously extend all the $g_{n,k}$'s to a homeomorphism g of class C^1 on $[\beta_0/2, 1]$, so that $g'|_{K_J} \equiv 2$.

Finally, let Q be $J \times J$ and $\varphi: Q \to Q$ be a diffeomorphism given by

$$\varphi(x,y) = (g(x), g^{-1}(y)), \qquad \text{if} \quad x \in \left[\frac{\beta_0}{2}, 1\right]$$

$$\varphi(x,y) \notin Q, \qquad \text{if} \quad |x| < \frac{\beta_0}{2}$$

$$\varphi(x,y) = (g(-x), -g^{-1}(y)), \qquad \text{if} \quad x \in \left[-1, -\frac{\beta_0}{2}\right].$$



graph of g, partially defined

The reader may easily verify that φ is of class C^1, $p = (1, -1)$ is a hyperbolic fixed point of φ, $\Lambda = \bigcap_{n \in \mathbf{Z}} \varphi^n(Q) = K_J \times K_J$ is a hyperbolic horseshoe and that $W^s_{loc}(p) \cap \Lambda = K_J$. Besides, both Λ and K_J have Lebesgue measure greater than zero.

§3. Local invariants and continuity

We conclude this chapter with some remarks on localized versions of the numerical invariants for Cantor sets introduced so far, and on the (continuous) dependence of these invariants on the Cantor set, at least for dynamically defined Cantor sets.

We give the definition of local thickness; local denseness, local Hausdorff dimension and local limit capacity are similarly defined. Let $K \subset \mathbf{R}$ be a Cantor set and $k \in K$. The local thickness $\tau_{loc}(K, k)$ of K at k is defined as

$$\tau_{loc}(K,k) = \lim_{\varepsilon \to 0} \sup\{\tau(\tilde{K}) \mid \tilde{K} \text{ is the intersection of } K \text{ with an interval,}$$

$$\text{contained in an } \varepsilon\text{-neighbourhood of } k\}.$$

For dynamically defined Cantor sets these notions have some additional properties. Let K be a dynamically defined Cantor set with expanding map Ψ. Then for every $U \subset K$, U open, there is some n so that $\Psi^n(U) = K$. From this and the bounded distortion property it follows that the local invariants $\tau_{loc}(K,k)$, $\theta_{loc}(K,k)$, $HD_{loc}(K,k)$, and $d_{loc}(K,k)$ are, in the dynamically defined case, all independent of k. Also, since the limit capacity and the Hausdorff dimension are invariant under diffeomorphisms, one has in this case $HD_{loc}(K,k) = HD(K) = d(K) = d_{loc}(K,k)$. The thickness and the denseness are not invariant under diffeomorphisms, and we may have $\tau(K) < \tau_{loc}(K,k)$ or $\theta(K) < \theta_{loc}(K,k)$.

For a discussion of the continuous dependence of the invariants on the Cantor set, we restrict ourselves to the dynamically defined case. ***Bearing in mind the dynamics of basic sets of surface diffeomorphisms, we define when two Cantor sets are near each other as follows.*** Let K be a Cantor set with expanding map Ψ and Markov partition $K_1, \ldots, K_\ell$. Suppose that Ψ is $C^{1+\varepsilon}$ with Hölder constant C, i.e. with $|\Psi'(p) - \Psi'(q)| \le C|p-q|^\varepsilon$ for all p, q in a neighbourhood of K. We say that the Cantor set $\tilde{K}$ is near K if $\tilde{K}$ has expanding map $\tilde{\Psi}$ and Markov partition $\tilde{K}_1, \ldots, \tilde{K}_\ell$ such that:

- $\tilde{\Psi}$ is $C^{1+\tilde{\varepsilon}}$ and is C^1 near Ψ; its derivative $\tilde{\Psi}'$ has Hölder constant $\tilde{C}$ such that $(\tilde{\varepsilon}, \tilde{C})$ is near (ε, C).
- $(\tilde{K}_1, \ldots, \tilde{K}_\ell)$ is near $(K_1, \ldots, K_\ell)$ in the sense that corresponding endpoints are near.

An important consequence of this definition is the existence, for nearby Cantor sets K and $\tilde{K}$ as above, of a homeomorphism $h : K \to \tilde{K}$, C^0-close to the identity, such that $\tilde{\Psi} \circ h = h \circ \Psi$. We construct h as follows. Notice first that, because of the proximity assumptions in the definition, $\Psi(K_i)$ intersects (and then contains) K_j if and only if the

same happens with $\tilde{\Psi}(\tilde{K}_i)$ and $\tilde{K}_j$. It follows that, given $x \in K$, there is $\tilde{x} \in \tilde{K}$ such that $\tilde{\Psi}^n(\tilde{x}) \in \tilde{K}_i \Leftrightarrow \Psi^n(x) \in K_i$, for all $n \ge 0$. Since $\tilde{\Psi}$ is expanding, $\tilde{x}$ must be unique; we define $h(x) = \tilde{x}$. Clearly $\tilde{\Psi}(h(x)) = h(\Psi(x))$. On the other hand we can obtain h^{-1} by a symmetrical construction, so h is really a bijection. Checking that h is close to the identity presents no particular difficulty. Just construct Markov partitions $\mathcal{R}^n$ and $\tilde{\mathcal{R}}^n$ for K and $\tilde{K}$ as in the previous section taking connected components of the inverse images of the K_j, respectively $\tilde{K}_j$, by Ψ^{n-1}, respectively $\tilde{\Psi}^{n-1}$. Then, one observes that x and $h(x)$ belong to corresponding intervals of $\mathcal{R}^n$ and $\tilde{\mathcal{R}}^n$ for all n and that corresponding intervals are uniformly (meaning independently of n) close, due to the closeness of $\tilde{K}$ to K, Ψ to $\tilde{\Psi}$ and to the bounded distortion property. We are left to show that h is continuous. We do more than that: we prove that it is Hölder continuous. Take $\delta > 0$ such that $d(K_i, K_j) > 3\delta$ and $d(\tilde{K}_i, \tilde{K}_j) > 3\delta$ for all $i \ne j$. Now, for $x, y \in K$ with $|x - y| \le \delta$ we let $n = n(x,y) \ge 0$ be such that

$$\left|\Psi^i(x) - \Psi^i(y)\right| \le 2\delta \text{ for } 0 \le i \le n-1$$

and

$$|\Psi^n(x) - \Psi^n(y)| \ge 2\delta.$$

By the definition of δ, the interval $[\Psi^i(x), \Psi^i(y)]$ is contained in some element of the Markov partition, for every $0 \le i \le n-1$. On the other hand we may assume that

$$\left|\tilde{\Psi}^i(\tilde{x}) - \tilde{\Psi}^i(\tilde{y})\right| \le 3\delta \text{ for } 0 \le i \le n-1.$$

To have this we just take $\tilde{K}$ close enough to K, in order to have $|h(x) - x| \le \delta/2$ for all x (note that $\tilde{\Psi}^i(\tilde{x}) = h(\Psi^i(x))$). Then again $[\tilde{\Psi}^i(\tilde{x}), \tilde{\Psi}^i(\tilde{y})]$ must be contained in some $\tilde{K}_j$, for all $0 \le i \le n-1$. By the mean value theorem there are $\xi_i \in [\Psi^i(x), \Psi^i(y)]$, $\tilde{\xi}_i \in [\tilde{\Psi}^i(\tilde{x}), \tilde{\Psi}^i(\tilde{y})]$ such that

$$|\Psi^n(x) - \Psi^n(y)| = |x - y| \prod_0^{n-1} |\Psi'(\xi_i)|$$

$$\left|\tilde{\Psi}^n(\tilde{x}) - \tilde{\Psi}^n(\tilde{y})\right| = |\tilde{x} - \tilde{y}| \prod_0^{n-1} \left|\tilde{\Psi}'(\tilde{\xi}_i)\right|.$$

Clearly, we can take $0 < \gamma < 1$ such that $|\Psi'(\xi_i)|^\gamma \leq \left|\tilde{\Psi}'(\tilde{\xi}_i)\right|$. Then we get

$$\frac{|\tilde{x} - \tilde{y}|}{|x - y|^\gamma} \leq \frac{\left|\tilde{\Psi}^n(\tilde{x}) - \tilde{\Psi}^n(\tilde{y})\right|}{|\Psi^n(x) - \Psi^n(y)|^\gamma} \leq (\text{ diam } K) \cdot \delta^{-\gamma}.$$

Hence h is Hölder continuous, as we claimed.

In fact we have proven even more. Since $\tilde{\Psi}$ is C^1-close to Ψ and $\tilde{\xi}_i$ is close to ξ_i, the values of $|\Psi'(\xi_i)|$ and $\left|\tilde{\Psi}'(\tilde{\xi}_i)\right|$ are almost equal. Therefore, if $\tilde{K}$ is close to K, *the Hölder exponent γ of the conjugacy h may be taken close to* 1. This, together with the analogous fact for h^{-1}, implies the following important result.

Theorem. *The Hausdorff dimension and limit capacity of a dynamically defined Cantor set K depend continuously on K.*

To derive the theorem from the considerations above one just has to observe that the existence of a homeomorphism $h: K \to \tilde{K}$ such that h and h^{-1} both are C^γ implies that $\gamma \cdot HD(K) \leq HD(\tilde{K}) \leq \gamma^{-1} \cdot HD(K)$ (and analogously for limit capacity). This, on its turn, is a direct consequence of the definitions.

Now we state and prove the corresponding result for thickness and denseness.

Theorem. *The thickness and the denseness of a dynamically defined Cantor set K depend continuously on K. The same holds for local thickness and local denseness.*

Heuristically, the theorem is proved as follows. The global strategy is to show that the values $\tau(K, \mathcal{U}, u)$, with $\mathcal{U} = \{U_n\}$ a presentation of K and u in the boundary of some bounded gap $U = U_n$, depend equicontinuously on K in the sense that if $\tilde{K}$ is close to K then $\tau(\tilde{K}, h(\mathcal{U}), h(u))$ is close to $\tau(K, \mathcal{U}, u)$ for all $\mathcal{U}$ and u. Here $h: K \to \tilde{K}$ is the conjugacy from Ψ to $\tilde{\Psi}$ described above (assume $\tilde{K}$ close enough to K to ensure that h exists) and $h(\mathcal{U})$ is the presentation of $\tilde{K}$ given by $h(\mathcal{U}) = \{h(U_n)\}$, where $h(U_n)$ is defined by $\partial h(U_n) = h(\partial U_n)$. Observe that h, as we constructed it, is monotonous.

For any given u and $\mathcal{U}$ we can, just by forcing h to be close enough to the identity, make $\tau(\tilde{K}, h(\mathcal{U}), h(u))$ arbitrarily close to $\tau(K, \mathcal{U}, u)$. We can even make this happen

simultaneously for all u (and $\mathcal{U}$) for which the corresponding gap U is big, say with length bigger than some fixed $\alpha > 0$. However such a simple argument is insufficient to obtain the uniform closeness that we need. To deal with the small gaps we must use the bounded distortion property. The idea is to iterate the gap U and the $\mathcal{U}$-component C of u until they become big. To be precise we fix $\beta > 0$ and take $k = k(U, C) \geq 0$ minimal such that $\ell(\Psi^k(U \cup C)) \geq \beta$. From the bounded distortion property we conclude that $\tau(K, \mathcal{U}, u) = \dfrac{\ell(C)}{\ell(U)}$ is almost equal to $\dfrac{\ell(\Psi^n(C))}{\ell(\Psi^n(U))}$: their ratio admits a bound depending only on $\beta > 0$ and K and which can be made arbitrarily close to 1 by taking β small enough. Analogously, from the bounded distortion property for $\tilde{K}, \tilde{\Psi}$ we obtain that $\tau(\tilde{K}, h(\mathcal{U}), h(u)) = \dfrac{\ell(h(C))}{\ell(h(U))}$ is almost equal to $\dfrac{\ell(\tilde{\Psi}^n(h(C)))}{\ell(\tilde{\Psi}^n(h(U)))}$. Moreover, and this is a key point, the bound for the ratio of this last two values may be taken to be independent of $\tilde{K}$ in a neighbourhood of K. This is a consequence of the fact that bounds for the distortion may be taken to be uniform in a neighbourhood of any Cantor set. To explain this, let us first observe that the positive numbers $c(\delta)$ constructed in the proof of the bounded distortion property vary continously with the dynamically defined Cantor set. In fact, these $c(\delta)$ depend only on the Hölder constants of the derivative of the expanding map and, by definition, nearby Cantor sets have expanding maps whose derivatives have nearby Hölder constants. In particular, it follows that we can take (new) upper bounds $c(\delta)$ as in the statement of the bounded distortion property which are uniform, i.e. independent of the Cantor set in a neighbourhood of K. We assume in what follows that $\tilde{K}$ belongs to this neighbourhood.

Now, if $\ell(\Psi^k(U))$ is big, that is larger than α, we can argue as before, i.e. use the proximity of h to the identity to conclude that $\dfrac{\ell(\tilde{\Psi}^k(h(C)))}{\ell(\tilde{\Psi}^k(h(C)))} = \dfrac{\ell(h(\Psi^k(C)))}{\ell(h(\Psi^k(U)))}$ is close to $\dfrac{\ell(\Psi^k(C))}{\ell(\Psi^k(U))}$. This, together with the estimates obtained above with the aid of the bounded distortion property, proves that $\tau(\tilde{K}, h(\mathcal{U}), h(u))$ is close to $\tau(K, \mathcal{U}, u)$, as we wanted to show.

Of course, we still have the problem that $\Psi^k(U)$ may be small. Iterating further is no solution, it may not be possible to do it, if $\Psi^j(C)$ gets out of the domain of Ψ before $\Psi^j(U)$ gets large. Even if this does not happen, as we iterate the length of $\Psi^j(U\cup C)$ gets bigger and so the bounds given by the bounded distortion property get rougher. Clearly, for the preceding argument we needed these bounds to be close to 1. Instead, what we do is to show that for our purposes this situation doesn't need to be taken into consideration. First, we observe that since $\ell(\Psi^k(U))\le\alpha$ and $\ell(\Psi^k(U\cup C))\ge\beta$, if we have chosen from the beginning $\beta>>\alpha$, then $\frac{\ell(\Psi^k(C))}{\ell(\Psi^k(U))}$ must be very big. The conjugacy h being close to the identity, the same holds for $\frac{\ell(\tilde\Psi^k(h(C)))}{\ell(\tilde\Psi^k(h(U)))}$. Using the bounded distortion property as above we conclude that $\tau(K,\mathcal{U},u)$ and $\tau(\tilde K,h(\mathcal{U}),h(u))$ are very big. Since in the calculation of both the thickness and the denseness one must at some point take an infimum, these values are irrelevant for this calculation and so may be disregarded when proving the continuity of $\theta(K)$ and $\tau(K)$.

We now come to a formal proof.

Proof: Let $A=\sup|\Psi'|$ and $B=2\theta(K)+8$. Let $\varepsilon>0$, $\delta>0$ and $\alpha>0$. Suppose that $\tilde K$ is close enough to K so that $|h(x)-x|\le\alpha\delta$ for all $x\in K$. We prove that if $\alpha>0$ and $\delta>0$ are chosen appropriately small (the precise conditions are given below) then this implies

$$(a)\quad \theta(\tilde K)\le(1+\varepsilon)^2\theta(K)+\varepsilon(1+\varepsilon)\qquad (b)\quad \theta(\tilde K)\ge(1+\varepsilon)^{-2}\theta(K)-\varepsilon(1+\varepsilon)^{-1}.$$

$$(c)\quad \tau(\tilde K)\le(1+\varepsilon)^2\tau(K)+\varepsilon(1+\varepsilon)\qquad (d)\quad \tau(\tilde K)\ge(1+\varepsilon)^{-2}\tau(K)-\varepsilon(1+\varepsilon)^{-1}$$

This proves the first part of the theorem. Then we show that the second part is an easy consequence of the first one.

First we take $\alpha>0$ small enough so that the $2AB\alpha$-neighbourhood of K is contained in the domain of Ψ. Clearly, we may assume that the same holds for $\tilde K$ and $\tilde\Psi$. For $\mathcal{U}=\{U_n\}$ a presentation of K, u a boundary point of a bounded gap $U=U_n$ and C the $\mathcal{U}$ component of K at u, take $k\ge0$ minimal such that $\ell(\Psi^k(U\cup C))\ge B\alpha$. Then $\ell(\Psi^k(U\cup C))\le AB\alpha$ (because $\ell(\Psi^{k-1}(U\cup C))\le B\alpha$) and so $\ell(\tilde\Psi^k(h(U)\cup h(C)))\le AB\alpha+2\alpha\delta\le 2AB\alpha$ (as

long as $\delta\le\frac{AB}{2}$). By the bounded distortion property we have

$$e^{-c(AB\alpha)}\le\left[\left(\frac{\ell(\Psi^k(C))}{\ell(\Psi^k(U))}\right)\Big/\left(\frac{\ell(C)}{\ell(U)}\right)\right]\le e^{c(AB\alpha)}\tag{1}$$

and

$$e^{-c(2AB\alpha)}\le\left[\left(\frac{\ell(\tilde\Psi^k(h(C)))}{\ell(\tilde\Psi^k(h(U)))}\right)\Big/\frac{\ell(h(C))}{\ell(h(U))}\right]\le e^{c(2AB\alpha)}\tag{2}$$

where $c(\ )$ is the distortion bounding function that we recalled in the heuristic proof. We assume that α is small enough so that this implies

$$(1+\varepsilon)^{-1}\le\left[\left(\frac{\ell(\Psi^k(C))}{\ell(\Psi^k(U))}\right)\Big/\left(\frac{\ell(C)}{\ell(U)}\right)\right]\le(1+\varepsilon)\tag{1a}$$

$$(1+\varepsilon)^{-1}\le\left[\left(\frac{\ell(\tilde\Psi^k(h(C)))}{\ell(\tilde\Psi^k(h(U)))}\right)\Big/\left(\frac{\ell(h(C))}{\ell(h(U))}\right)\right]\le(1+\varepsilon)\tag{2a}$$

Now we distinguish two cases according to the size of Ψ. Suppose first that $\ell(\Psi^k(U))\ge\alpha$. Then

$$\left|\frac{\ell(\Psi^k(C))}{\ell(\Psi^k(U))}-\frac{\ell(h(\Psi^k(C)))}{\ell(h(\Psi^k(U)))}\right|\le$$

$$\le\frac{\ell(\Psi^k(C))\cdot|\ell(h(\Psi^k(U)))-\ell(\Psi^k(U))|+\ell(\Psi^k(U))\cdot|\ell(h(\Psi^k(C)))-\ell(\Psi^k(C))|}{\ell(\Psi^k(U))\cdot\ell(h(\Psi^k(U)))}\le$$

$$\le\frac{AB\alpha\cdot2\alpha\delta+AB\alpha\cdot2\alpha\delta}{\alpha\cdot(\alpha-2\alpha\delta)}=\delta\frac{4AB}{1-2\delta}\tag{3}$$

If $\delta>0$ is sufficiently small this implies

$$\left|\frac{\ell(\Psi^k(C))}{\ell(\Psi^k(U))}-\frac{\ell(\tilde\Psi^k(h(C)))}{\ell(\tilde\Psi^k(h(U)))}\right|\le\varepsilon.\tag{3a}$$

From (1a), (2a) and (3a) it immediately follows that

$$\tau(\tilde K,h(\mathcal{U}),h(u))\le(1+\varepsilon)\cdot((1+\varepsilon)\cdot\tau(K,\mathcal{U},u)+\varepsilon)\tag{4a}$$

and

$$\tau(\tilde K, h(\mathcal{U}), h(u)) \geq (1+\varepsilon)^{-1}((1+\varepsilon)^{-1}\tau(K,\mathcal{U},u)-\varepsilon). \tag{4b}$$

Let now $\ell(\Psi^k(U)) \leq a$. Then we must have $\ell(\Psi^k(C)) \geq Ba - a$. Moreover $\ell(\tilde\Psi^k(h(U))) \leq a + 2a\delta$ and $\ell(\tilde\Psi^k(h(C))) \geq Ba - a - 2a\delta$. This together with (1) and (2) implies

$$\frac{\ell(C)}{\ell(U)} \geq e^{-c(ABa)} \cdot \frac{Ba - a}{a} = (B-1)e^{-c(ABa)} \tag{5}$$

and

$$\frac{\ell(h(C))}{\ell(h(U))} \geq e^{-c(2ABa)} \cdot \frac{Ba - 2a\delta - a}{a + 2a\delta} = \frac{B - 2\delta - 1}{1 + 2\delta} \cdot e^{-c(2ABa)} \tag{6}$$

Since we have chosen $B = 2\theta(K) + 8$ we can suppose a and δ small enough so that this implies

$$\tau(K,\mathcal{U},u) \geq (\theta(K)+3) \tag{5a}$$

and

$$\tau(\tilde K, h(\mathcal{U}), h(u)) \geq (\theta(K)+3). \tag{6a}$$

Now we proceed to prove the affirmatives (a)-(d) stated near the beginning of the proof. Recall that by definition

$$\tau(K) = \sup_{\mathcal{U}} \inf_{u} \quad \tau(K,\mathcal{U},u)$$

$$\theta(K) = \inf_{\mathcal{U}} \sup_{u} \quad \tau(K,\mathcal{U},u).$$

To prove (a) we must find for any given $\mathcal{U}$ a presentation $\tilde{\mathcal{U}}$ of $\tilde K$ such that

$$\sup_{\tilde u} \tau(\tilde K,\tilde{\mathcal{U}},\tilde u) \leq (1+\varepsilon)^2 \sup_{u} \tau(K,\mathcal{U},u) + \varepsilon(1+\varepsilon) \tag{a1}$$

We loose no generality in assuming that

$$\sup_{u} \tau(K,\mathcal{U},u) \leq \theta(K) + 1. \tag{a2}$$

Take $\tilde{\mathcal{U}} = h(\mathcal{U})$. From (a2) it follows that $\tau(K,\mathcal{U},u) \leq \theta(K)+1$ for all u and so (5a) never holds. Then, by the previous discussion we must have (4a)

$$\tau(\tilde K,\tilde{\mathcal{U}},\tilde u = h(u)) \leq (1+\varepsilon)^2 \cdot \tau(K,\mathcal{U},u) + \varepsilon(1+\varepsilon) \tag{4a}$$

(as well as (4b)) for all u. This immediately implies (a1) and so (a) is proved.

The proof of (b) is almost dual to the preceding one so we don't write it down in detail. The only asymmetry comes from the fact that (6a) involves $\theta(K)$ and not $\theta(\tilde K)$. This is bypassed as follows. First, we may as above suppose that

$$\sup_{\tilde u} \tau(\tilde K,\tilde{\mathcal{U}},\tilde u) \leq \theta(\tilde K) + 1. \tag{b2}$$

Now, from (a) (that we have already proved) we get that if $\tilde K$ is sufficiently near K then $\theta(\tilde K) \leq \theta(K) + 1$. Then (b2) implies

$$\sup_{\tilde u} \tau(\tilde K,\tilde{\mathcal{U}},\tilde u) \leq \theta(K) + 2 \tag{b3}$$

and now the argument proceeds as before.

To prove (c) we take, for each $\tilde{\mathcal{U}}$, $\mathcal{U} = h^{-1}(\tilde{\mathcal{U}})$ and show that

$$\inf_{\tilde u} \tau(\tilde K,\tilde{\mathcal{U}},\tilde u) \leq (1+\varepsilon)^2 \inf_{u} \tau(K,\mathcal{U},u) + \varepsilon(1+\varepsilon). \tag{c1}$$

To do this we must associate to each u and $\tilde u$ such that

$$\tau(\tilde K,\tilde{\mathcal{U}},\tilde u) \leq (1+\varepsilon)^2 \cdot \tau(K,\mathcal{U},u) + \varepsilon(1+\varepsilon). \tag{c2}$$

Again, it is sufficient to consider the points u for which

$$\tau(K,\mathcal{U},u) \leq \inf_{u} \tau(K,\mathcal{U},u) + 1. \tag{c3}$$

Take $\tilde u = h(u)$ and observe that if (c3) holds then $\tau(K,\mathcal{U},u) \leq \tau(K) + 1 \leq \theta(K) + 1$ and so (5a) doesn't hold. Therefore (4a) is true, and this is just (c2). The proof of (c) is complete.

The proof of (d) is dual to the one of (c) (recall also the remark in the proof of (b)) so we are done with proving the continuity of (global) thickness and denseness.

Finally, recall that the local thickness of a Cantor set K at a point $k \in K$ is defined by

$$\tau_{\text{loc}}(K,k) = \lim_{\delta \to 0}(\sup\{\tau(K_1) | K_1 \subset K \cap B_\delta(k), \text{ a Cantor set }\}).$$

Let $\varepsilon > 0$ be small. Given $\tilde{\delta} > 0$, take $\delta > 0$ such that $h(K \cap B_\delta(k)) \subset \tilde{K} \cap B_{\tilde{\delta}}(h(k))$. Let K_1 be a Cantor set in $K \cap B_\delta(k)$ and let $\tilde{K}_1 = h(K_1)$. If h is close enough to the identity (i.e. if $\tilde{K}$ is close enough to K) then the arguments above imply $\tau(\tilde{K}_1) \geq \tau(K_1) - \varepsilon$. Since K_1 is arbitrary it follows that

$$\sup\{\tau(K_1) | K_1 \subset K \cap B_\delta(k), \text{ a Cantor set }\} \leq$$
$$\leq \sup\{\tau(\tilde{K}_1) | \tilde{K}_1 \subset \tilde{K} \cap B_{\tilde{\delta}}(h(k)), \text{ a Cantor set }\} + \varepsilon.$$

By making $\tilde{\delta} \to 0$ (and so $\delta \to 0$) we get

$$\tau_{\text{loc}}(K,k) \leq \tau_{\text{loc}}(\tilde{K}, h(k)) + \varepsilon.$$

In the same way one shows

$$\tau_{\text{loc}}(K,k) \geq \tau_{\text{loc}}(\tilde{K}, h(k)) - \varepsilon.$$

This shows the continuity of local thickness. For local denseness the argument is the same. The proof of the theorem is now complete. ∎

Remark: Consider a C^3 diffeomorphism φ of a surface, with a basic set Λ and a saddle point $p \in \Lambda$. For $\tilde{\varphi}$ a C^3 nearby diffeomorphism there are $\tilde{\Lambda}$, a basic set, and $\tilde{p} \in \tilde{\Lambda}$, a saddle point (near Λ and p, respectively), and the dynamically defined Cantor sets $W^u(p) \cap \Lambda$ and $W^u(\tilde{p}) \cap \tilde{\Lambda}$ are near in the above sense (if we take nearby parametrizations for $W^u(p)$ and $W^u(\tilde{p})$ as in Section 1 of this present chapter). This follows from the continuous dependence on the diffeomorphism of basic sets and their $C^{1+\varepsilon}$ stable and unstable foliations; see Appendix I and Remark 2 in Appendix II, concerning continuous dependence of Markov partitions. From this and the propositions that we just proved, we

conclude the continuous dependence with respect to the diffeomorphism in the C^3 topology, of all the invariants of $W^u(p) \cap \Lambda$ that we have discussed, namely Hausdorff dimension, limit capacity, thickness and denseness. To show this, one uses the arguments above together with the observation that C^3 diffeomorphisms and C^3 closeness are used only to obtain $C^{1+\varepsilon}$ expanding maps with nearby Hölder constants for the derivatives. This in turn provides bounds for the distortion of distances which are uniform in neighbourhoods of the diffeomorphism and the Cantor set. But, as we remarked before, at the end of Section 1, C^2 diffeomorphisms induce Cantor sets satisfying the bounded distortion property (and the resulting expanding maps are indeed $C^{1+\varepsilon}$ for some $\varepsilon > 0$). The argument that we used there also yields uniform estimates for the distortion in a C^2 neighbourhood of the original diffeomorphism. *Thus, all the above invariants of $W^u(p) \cap \Lambda$ depend continuously on φ in the C^2 topology.*

For Hausdorff dimension and limit capacity, one can go even further: in **[PV, 1988]** it is proved that the ***Hausdorff dimension and limit capacity*** of $W^u(p) \cap \Lambda$ ***depend continuously on the diffeomorphism in the C^1 topology.*** This is done by using, as above, conjugacies with Hölder constants near 1. We observe that this result had been obtained in [MM,1983] as a consequence of a variational principle of the thermodynamical formalism.

REFERENCES

[A,1961] V.I. Arnold – On the mapping of the circle into itself, Izvestia Akad. Nauk. Math. Series bf **25** (1961), 21-86.

[A,1961] V.I. Arnold – Small denominators and problems of stability of motion in classical and celestial mechanics, Russ. Math. Surveys **18**, **(6)** (1963), 85-192.

[A,1980] V.I. Arnold – Chapitres suplémentaires de la théorie des équations différentielles ordinaires, MIR, Moscow, 1980.

[A,1967] D.V. Anosov – Geodesic flows on closed Riemannian manifolds with negative curvature, Proc. Stek. Inst. **90** (1967), A.M.S transl. (1969).

[A,1991] N. Aoki – The set of Axiom A diffeomorphisms with no cycles, Bol. Soc. Bras. Mat., to appear.

[AM,1991] A. Araujo and R. Mañé – On the existence of hyperbolic attractors and homoclinic tangencies for surface diffeomorphisms, to appear.

[AP,1935] A. Andronov and L. Pontryagin – Systèmes grossiers, Dokl. Akad. Nauk. USSR **14** (1937), 247-251.

[AP,1987] V.S. Afraimovich and Ya.B. Pesin – Mathematical Physics Reviews, Section C, **6** (1987), 169-241.

[B,1935] G.D. Birkhoff – Nouvelles recherches sur les systèmes dynamiques, Mem. Pont. Acad. Sci. Novi. Lyncaei **1** (1935), 85-216.

[B,1975] R. Bowen – Equilibrium states and the ergodic theory of Anosov diffeomorphisms, Lecture Notes in Math. **470** (1975), Springer-Verlag.

[B,1975] R. Bowen – A horseshoe with positive measure, Inventiones Math. **29** (1975), 203-204.

[B,1977] R. Bowen – On Axiom A diffeomorphisms, Conference Board Math. Sci. **33**, A.M.S, 1977.

[BR,1975] R. Bowen, D. Ruelle – The ergodic theory of Axiom A flows, Inventiones Math **29** (1975), 181-202.

[BC,1985] M. Benedicks and L. Carleson – On iterations of $1 - ax^2$ on $(-1,1)$, Annals of Math. **122** (1985), 1-24.

[BC,1991] M. Benedicks and L. Carleson – The dynamics of the Hénon map, Annals of Math. **133** (1991), 73-169.

[BY,1991] M. Benedicks and L.S. Young – SBR measures for certain Hénon maps, to appear.

[BF,1973] L. Block and J. Franke – Existence of periodic points for maps of S^1, Inventiones Math. **22** (1973), 69-73.

[BHTB,1990] H.W. Broer, G.B. Huitema, F. Takens and B.L.J. Braskma – Unfoldings and bifurcations of quasi-periodic tori, Memoirs of the A.M.S. **83**, **n. 421** (1990).

[BLMP,1991] R. Banón, R. Labarca, R. Mañé and M.J. Pacífico – Bifurcating 3-dimensional singular cycle, to appear.

[BS,1979] A.A. Bunimovich and Ya.G. Sinai – Stochastic attractors in the Lorenz model, Nonlinear Waves, Nauka (1979), 212-226.

[C,1991] M. Carvalho – Bowen-Ruelle-Sinai measures for n–dimensional DA diffeomorphisms, IMPA's thesis and to appear.

[C,1978] C. Conley – Isolated invariant sets and the Morse index, Conference Board Math. Sci. **38**, A.M.S, 1978.

[CE,1980] P. Collet, J.-P. Eckmann – Iterated maps on the interval as dynamical systems, Birkhäuser, 1980.

[CL,1945] M.L. Cartwright and J.E. Littlewood – On nonlinear differential equations of the second order: I. The equation $\ddot{y} - k(1 - y^2)\dot{y} + y = b\lambda k\cos(\lambda t + a)$, k large, J. London Math. Soc. (1945), 180-189.

[D,1991] L. Diaz – Persistence of nonhyperbolicity and heterodimensional cycles, IMPA's thesis and to appear.

[DR,1991] L. Diaz and J. Rocha - Nonconnected heterodimensional cycles: bifurcations and stability, to appear.

[DRV,1991] L. Diaz, J. Rocha and M. Viana - Saddle-node critical cycles and prevalence of strange attractors, to appear.

[ER,1985] J.-P. Eckmann, D. Ruelle - Ergodic theory of chaos and strange attractors, Rev. Mod. Phys. **57** (1985), 617-656.

[F,1985] K. J. Falconer - The geometry of fractal sets, Cambridge Univ. Press, 1985.

[F,1985] J. Franks - Period doubling and the Lefschetz formula, Trans. A.M.S. 287 (1985), 275-283.

[FS,1977] J. Franke, J. Selgrade - Hyperbolicity and chain recurrence, Trans. A.M.S. **245** (1978), 251-262.

[GH,1983] J. Guckenheimer, P. Holmes - Nonlinear oscillations, dynamical systems and bifurcations of vector fields, Springer-Verlag, 1983.

[GW,1979] J. Guckenheimer and R. R. Williams - Structural stability of Lorenz attractors, Publ. Math. I.H.E.S. **50** (1979), 59-72.

[GST,1989] J.-M. Gambaudo, S. von Strein and C. Tresser - Hénon-like maps with strange attractors: there exist C^∞ Kupka-Smale diffeomorphisms on S^2 with neither sinks or sources, Nonlinearity **2** (1989), 287-304.

[H,1947] M. Hall - On the sum and product of continued fractions, Annals of Math. 48 (1947), 966-993.

[H,1964] P. Hartman - Ordinary differential equations, Wiley, 1964.

[H,1976] M. Hénon - A two-dimensional mapping with a strange attractor, Comm. of Math. Phys. **50** (1976), 69-77.

[H,1977] M. Herman - Measure de Lebesgue et nombre de rotation, Geometry and Topology, ed. J. Palis and M. do Carmo, Lecture Notes in Math. **597** (1978), 271-293, Springer-Verlag.

[H,1979] M. Herman - Sur la conjugaison différentiable des difféomorphismes du cercle à des rotations, Publ. Math. **49** (1979), 5-233.

[H,1991] S. Hayashi - Diffeomorphisms in $\mathcal{F}^1(M)$ satisfy Axiom A, Erg. Th. and Dyn. Syst., to appear.

[HPPS,1970] M. Hirsch, J. Palis, C. Pugh and M. Shub - Neighbourhoods of hyperbolic sets, Inventiones Math. **9** (1970), 121-134.

[HPS,1977] M. Hirsch, C. Pugh and M. Shub - Invariant manifolds, Lecture Notes in Math. **583** (1977), Springer-Verlag.

[J,1971] M. Jacobson - On smooth mappings of the circle into itself, Math. USSR Sb. **14** (1971), 161-185.

[K,1957] A.N. Kolmogorov - General theory of dynamical systems and classical mechanics, Proc. Int. Congress of Math. 1954, 315-333, North Holland

[K,1964] I. Kupka - Contribution à la théorie des champs génériques, Cont. Diff. Equ. **2** (1963), 457-484, **3** (1964), 411-420.

[K,1980] A. Katok - Lyapunov exponents, entropy and periodic orbits for diffeomorphisms, Publ. Math. IHES **51** (1980), 137-174.

[L,1949] N. Levinson - A second order differential equation with singular solutions, Annals of Math. **50** (1949), 127-153.

[L,1957] J.E. Littlewood - On non-linear differential equations of second order: III, Acta Math. 97 (1957), 267-308.

[L,1957] J.E. Littlewood - On non-linear differential equations of second order: IV, Acta Math. **98** (1957), 1-110.

[L,1963] E.N. Lorenz - Deterministic non-periodic flow, J. Atmos. Sci. **20** (1963), 130-141.

[L,1980] S.T. Liao - On the stability conjecture, Chinese Annals of Math. **1** (1980), 9-30.

[L,1981] M. Levi - Qualitative analysis of the periodically forced relaxation oscillations, Memoirs of the A.M.S. **32**, n. **244** (1981).

[LP,1986] R. Labarca and M.J. Pacifico - Stability of singular horseshoe, Topology 25 (1986), 337-352.

[L,1990] P. Larsson - L'ensemble différence de deux ensembles de Cantor aléatoires, C.R.A.S. Paris 310 (1990), 735-738.

[M,1954] J.M. Marstrand - Some fundamental properties of plane sets of fractional dimensions, Proc. London Math. Soc. 4 (1954), 257-302.

[M,1962] J. Moser - On invariant curves of area preserving mappings of an annulus, Nachr. Akad. Wiss. Göttingen, Math. Phys. Kl (1962), 1-20.

[M,1967] J. Moser - Convergent series expansions for quasi-periodic motions, Math. Annalen. 169 (1967), 136-176.

[M,1969] J. Moser - On a theorem of Anosov, J. Diff. Equ. 5 (1969), 411-440.

[M,1973] W. de Melo - Structural stability of diffeomorphisms on two-manifolds, Inventiones Math. 21 (1973), 233-246.

[M,1978] R. Mañé - Contribution to the stability conjecture, Topology 17 (1978), 383-396.

[M,1982] R. Mañé - An ergodic closing lemma, Annals of Math. 116 (1982), 503-540.

[M,1987] R. Mañé - Ergodic theory and differentiable dynamics, Springer-Verlag, 1987.

[M,1988] R. Mañé - A proof of the C^1 stability conjecture, Publ. Math. I.H.E.S. 66 (1988), 161-210.

[MS,1980] M. Misiurewicz and B. Szewc - Existence of a homoclinic orbit for the Hénon map, Comm. Math. Phys. 75 (1980), 285-291.

[MM,1983] A. Manning and H. McCluskey - Hausdorff dimension for horseshoes, Erg. Th. and Dyn. Syst. 3 (1983), 251-261.

[MV,1991] L. Mora and M. Viana - Abundance of strange attractors, Acta Math., to appear.

[N,1970] S. Newhouse - Non-density of Axiom A(a) on S^2, Proc. A.M.S. Symp. Pure Math. 14 (1970), 191-202.

[N,1972] S. Newhouse - Hyperbolic limit sets, Trans. A.M.S. 167 (1972), 125-150.

[N,1974] S. Newhouse - Diffeomorphisms with infinitely many sinks, Topology 13 (1974), 9-18.

[N,1979] S. Newhouse - The abundance of wild hyperbolic sets and nonsmooth stable sets for diffeomorphisms, Publ. Math. I.H.E.S. 50 (1979), 101-151.

[N,1980] S. Newhouse - Lectures on dynamical systems, in: J. Guckenheimer, J. Moser, S. Newhouse, Dynamical systems, CIME Lectures - Bressanone, Birkhäuser, 1980.

[NP,1973] S. Newhouse and J. Palis - Hyperbolic nonwandering sets on two-manifolds, Dynamical systems, ed. M. Peixoto, Acad. Press, 1973, 293-301.

[NP,1976] S. Newhouse and J. Palis - Cycles and bifurcation theory, Astérisque 31 (1976), 44-140.

[NPT,1983] S. Newhouse, J. Palis and F. Takens - Bifurcations and stability of families of diffeomorphisms, Publ. Math. I.H.E.S. 57 (1983), 5-71.

[O,1968] V. Oseledec - A multiplicative ergodic theorem: Lyapunov characteristic numbers for dynamical systems, Trans. Moscow Math. Soc. 19 (1968), 197-231.

[P,1890] H. Poincaré - Sur le problème des trois corps et les équations de la dynamique (Memoir couronné du prise de S.M. le roi Oscar II de Suède), Acta Math. 13 (1890), 1-270.

[P,1920] B. van der Pol - De amplitude van vrije en gedwongen triode-trillingen, Tijdschr. Ned. Radiogenoot. 1 (1920), 3-31.

[P,1934] B. van der Pol - The nonlinear theory of electric oscillations, Proc. of the Inst. of Radio Eng. 22 (1934), 1051-1086; reprinted in: Selected scientific papers, North-Holland, 1960.

[P,1960] V. Pliss - Sur la grossièreté des equations différentielles définies sur le tore, Vestnik LGU, ser. mat. 13 (1960), 15-23.

[P,1962] M. Peixoto - Structural stability on two-manifolds, Topology **1** (1962), 101-120.

[P,1967] C. Pugh - The closing lemma, Amer. J. Math. **89** (1967), 956-1009.

[P,1969] J. Palis - On Morse Smale dynamical systems, Topology **8** (1969), 385-405.

[P,1970] J. Palis - A note on Ω-stability, Proc. A.M.S. Symp. Pure Math. **14** (1970), 221-222.

[P,1988] J. Palis, On the C^1 Ω-stability conjecture, Publ. Math. I.H.E.S. **66** (1988), 211-215.

[PM,1982] J. Palis and W. de Melo - Geometric theory of dynamical systems, Springer-Verlag, 1982.

[PS,1970] J. Palis and S. Smale - Structural stability theorems, Proc. A.M.S. Symp. Pure Math. **14** (1970), 223-232.

[PT,1985] J. Palis and F. Takens - Cycles and measure of bifurcation sets for two-dimensional diffeomorphims, Inventiones Math. **82** (1985), 397-422.

[PT,1987] J. Palis and F. Takens - Hyperbolicity and the creation of homoclinic orbits, Annals of Math. **125** (1987), 337-374.

[PV,1988] J. Palis and M. Viana - Continuity of Hausdorff dimension and limit capacity for horseshoes, Dynamical Systems, Lecture Notes in Math. **1331**, (1988), 150-160.

[PV,1988] J. Palis and M. Viana - Infinitely many sinks in higher dimensions, to appear.

[PY,1989] J. Palis and J. C. Yoccoz - Homoclinic tangencies for hyperbolic sets of large Hausdorff dimension, to appear.

[P,1974] R. V. Plykin Sources and currents of A-diffeomorphisms of surfaces, Math. Sbornik **94** (1974), 2, 243-264.

[P,1976] Ya. Pesin - Families of invariant manifolds corresponding to non-zero characteristic exponents, Math. of USSR Izvestjia **10** (1976), 1261-1305.

[P,1977] Ya. Pesin - Characteristic Lyapunov exponents and ergodic theory, Russian Math. Surveys **32** (1977), 55-114.

[P,1983] D. Pixton, Markov neighbourhoods for zero - dimensional basic sets, Trans. A.M.S. **279** (1983), 431-462.

[R,1986] M. Rees - Positive measure sets of ergodic rational maps, Ann. Scient. Éc. Norm. Sup. **19** (1986), 383-407.

[R,1971] J. Robbin - A structural stability theorem, Annals of Math. **93** (1971), 447-493.

[R,1973] C. Robinson - C^r structural stability implies Kupka-Smale, Dynamical Systems, ed. M. Peixoto, Acad. Press, 1973, 443-449.

[R,1974] C. Robinson - Structural stability of vector fields, Ann. Math **99** (1974), 154-175.

[R,1976] C. Robinson - Structural stability of C^1 diffeomorphisms, J. Diff. Equ. **22** (1976), 28-73.

[R,1983] C. Robinson - Bifurcation to infinitely many sinks, Comm. Math. Phys. **90** (1983), 433-459.

[R,1989] C. Robinson - Homoclinic bifurcation to a transitive attractor of Lorenz type, Nonlinearity **2** (1989), 495-518.

[R,1991] A. Rovella - The dynamics of the perturbations of the contracting Lorenz attractor, IMPA's thesis and to appear.

[R,1989] M. Rychlik - Lorenz's attractors through Silnikov-type bifurcation, Erg. Th. and Dyn. Syst. **10** (1990), 793-821.

[R,1976] D. Ruelle - A measure associated with Axiom A attractors, Amer. J. Math. **98** (1976), 619-654.

[R,1978] D. Ruelle - An inequality of the entropy of differentiable maps, Bol. Soc. Bras. Mat. **9** (1978), 83-87.

[RT,1971] D. Ruelle and F. Takens - Comm. Math. Phys. **20** (1971), 167-192 and **23** (1971), 343-344.

[S,1983] A. Sannami - The stability theorems for discrete dynamical systems on two-dimensional manifolds, Nagoya Math. J. **90** (1983), 1-55.

[S,1990] A. Sannami - An example of a regular Cantor set whose difference is a Cantor set with positive measure, Hokkaido Math. J., to appear.

[S,1971] M. Shub - Topological transitive diffeomorphisms on T^4, Lecture Notes in Math. **206** (1971), 39, Springer-Verlag.

[S,1978] M. Shub - Stabilité global des systèmes dynamiques, Astérisque **56** (1978).

[S,1965] L.P. Silnikov - A case of the existence of denumerable set of periodic motions, Sov. Math. Dokl. **6** (1965), 163-166.

[S,1976] Ya. Sinai - Introduction to ergodic theory, Princeton University Press, 1976.

[S,1968] Ya. Sinai - Markov partitions and C-diffeomorphisms, Func. Anal. and its Appl. **2** (1968), 64-89.

[S,1970] Ya. Sinai - Dynamical systems with elastic reflections: properties of dispersing billards, Russ. Math. Surveys **25** (1970), 137- 189.

[S,1963] S. Smale - Stable manifolds for differential equations and diffeomorphisms, Ann. Scuola Sup. Piza **17** (1963), 97-116.

[S,1965] S. Smale - Diffeomorphisms with many periodic points. Diff. and Comb. Topology, Princeton Univ. Press (1965), 63-80.

[S,1967] S. Smale - Differentiable dynamical systems, Bull. A.M.S. **73** (1967), 747-817.

[S,1970] S. Smale - The Ω-stability theorem, Proc. A.M.S. Symp. Pure Math. **14** (1970), 289-297.

[S,1957] S. Sternberg - Local contractions and a theorem of Poincaré, Amer. J. Math. **79** (1957), 809-824.

[S,1958] S. Sternberg - On the structure of local homeomorphisms of Euclidean n-space, II, Amer. J. Math. **80** (1958), 623-631.

[S,1979] S.J. van Strien - Centre manifolds are not C^∞, Math. Z. **166** (1979), 143-145.

[S,1981] S.J. van Strien - On the bifurcations creating horseshoes, Dynamical Systems, Lecture Notes in Math. **898** (1981), 316- 351, Springer-Verlag.

[SM,1971] C.L. Siegel and J. Moser - Lectures on celestial mechanics, Springer-Verlag, 1971.

[T,1988] F. Takens - Limit capacity and Hausdorff dimension of dynamically defined Cantor sets, Dynamical Systems, Lecture Notes in Math. **1331** (1988), 196-212, Springer-Verlag.

[T,1991] F. Takens - On the geometry of non-transversal intersections of invariant manifolds and scaling properties of bifurcation sets, Pitman Research Notes in Math. Series, to appear.

[T,1991] F. Takens - Homoclinic bifurcations, Proc. Int. Congress of Math., Berkeley (1986), 1229-1236.

[T,1991] F. Takens - Abundance of generic homoclinic tangencies in real-analytic diffeomorphisms, to appear.

[TY,1986] L. Tedeschini-Lalli and J.A. Yorke - How often do simple dynamical processes have infinitely many coexisting sinks? Comm. Math. Phys. **106** (1986), 635-657.

[V,1991] E. Vargas - Bifurcation frequency for unimodal maps, Comm. Math. Phys., to appear.

[V,1991] M. Viana - Strange attractors in higher dimensions, IMPA's thesis and to appear.

[W,1970] R.F. Williams - The DA maps of Smale and structural stability, Proc. A.M.S. Symp. Pure Math. **14** (1970), 329-334.

[W,1979] R.F. Williams - The structure of Lorenz attractors, Publ. Math. I.H.E.S. **50** (1979), 101-152.

[YA 1983] J. A. Yorke - K. T. Alligood, Cascades of period doubling bifurcations: a prerequisite for horseshoes, Bull. A.M.S. **9** (1983), 319-322.